**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68, C12P 19/34, C07H 21/00, 21/02, 21/04 — A1

(11) International Publication Number: **WO 00/32823**

(43) International Publication Date: 8 June 2000 (08.06.00)

(21) International Application Number: PCT/US99/28472

(22) International Filing Date: 2 December 1999 (02.12.99)

(30) Priority Data:
60/110,549 2 December 1998 (02.12.98) US

(71) Applicant: PHYLOS, INC. [US/US]; 128 Spring Street, Lexington, MA 02421 (US).

(72) Inventors: LOHSE, Peter; 50 Golden Ball Road, Weston, MA 02493 (US). KURZ, Markus; 62 Kensington Street, West Newton, MA 02460 (US). WAGNER, Richard; 1007 Lowell Road, Concord, MA 01742 (US).

(74) Agent: ELBING, Karen, L.; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110-2214 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

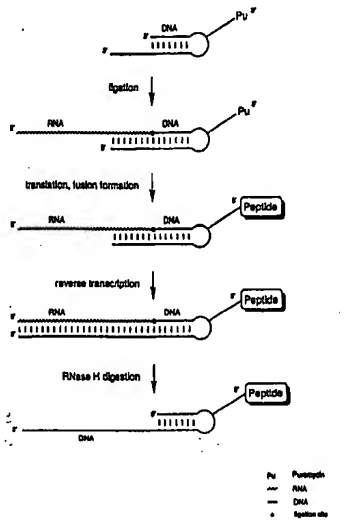
Published
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: DNA-PROTEIN FUSIONS AND USES THEREOF

(57) Abstract

The invention provides methods for covalently tagging proteins with their encoding DNA sequences as shown in the Figure. These DNA-protein fusions may be used in molecular evolution and recognition techniques.

DNA-Protein Fusion: Type A1




*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# DNA-PROTEIN FUSIONS AND USES THEREOF

## Background of the Invention

In general, the invention features DNA-protein fusions and their uses, particularly for the selection of desired proteins and their corresponding nucleic acid sequences.

Recently, a combinatorial method was developed for the isolation of proteins with desired properties from large pools of proteins (Szostak et al., U.S.S.N. 09/007,005; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302). By this method, the protein portion is linked to its encoding RNA by a covalent chemical bond. Due to the covalent nature of this linkage, selection experiments are not limited to the extremely mild reaction conditions that must be used for approaches that involve non-covalent complex formation such as ribosome display (Hanes & Plückthun, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 4937-4942; He & Taussig, Nucl. Acids Res. (1997) vol. 25, p 5132-5143). However, precautions do need to be taken during the selection process to minimize RNA degradation, since the accidental cleavage of ribo-bonds can result in the irreversible loss of encoded information. For this reason, these selection procedures are typically carried out using reaction media and equipment that are free of ribonucleases or other deleterious contaminants.

## Summary of the Invention

The present invention provides methods for covalently tagging proteins with their encoding DNA sequences. These DNA-protein fusions, which may be used in molecular evolution and recognition techniques, are chemically more stable than RNA-protein fusions and therefore provide a number of advantages (as discussed in more detail below).



Accordingly, in general, the invention features methods for generating DNA-protein fusions. A first method involves: (a) linking a nucleic acid primer to an RNA molecule (preferably, at or near the RNA 3' end) , the primer being bound to a peptide acceptor (for example, puromycin); (b) translating the RNA to produce a protein product, the protein product being covalently bound to the primer; and (c) reverse transcribing the RNA to produce a DNA-protein fusion.

A second method involves: (a) generating an RNA-protein fusion; (b) hybridizing a nucleic acid primer to the fusion (preferably, at or near the RNA 3' end); and (c) reverse transcribing the RNA to produce a DNA-protein fusion.

In a preferred embodiment of the above methods, the method may further involve treating the product of step (c) to remove the RNA (for example, by contacting the product of step (c) with RNase H under conditions sufficient to digest the RNA). In additional preferred embodiments, the nucleic acid primer is a DNA primer; the translating step is carried out in vitro; and the nucleic acid primer has a hairpin structure. In addition, the primer may further include a photocrosslinking agent, such as psoralen, and the primer may be crosslinked to an oligonucleotide which is bound to a peptide acceptor or, alternatively, may be hybridized to the RNA molecule, followed by a linking step that is carried out by photocrosslinking.

In related aspects, the invention also features a molecule including a DNA covalently bonded to a protein (preferably, of at least 10 amino acids) through a peptide acceptor (for example, puromycin), as well as a molecule including a DNA covalently bonded to a protein, in which the protein includes at least 10 amino acids.

In preferred embodiments of both of these aspects, the protein includes at least 30 amino acids, more preferably, at least 100 amino acids, and

may even include at least 200 or 250 amino acids. In other preferred embodiments, the protein is encoded by the DNA and is preferably entirely encoded by the DNA; the molecule further includes a ribonucleic acid covalently bonded to the DNA; the protein is encoded by the ribonucleic acid; and the DNA is double stranded.

In another related aspect, the invention features a population of at least $10^5$, and preferably, at least $10^{14}$, DNA-protein fusions of the invention, each fusion including a DNA covalently bonded to a protein.

In addition, the invention features selection methods which utilize the DNA-protein fusions described herein. A first selection method involves the steps of: (a) providing a population of DNA-protein fusions, each including a DNA covalently bonded to a candidate protein; and (b) selecting a desired DNA-protein fusion, thereby selecting the desired protein or DNA.

A second selection method involves the steps of: (a) producing a population of candidate DNA-protein fusions, each including a DNA covalently bonded to a candidate protein and having a candidate protein coding sequence which differs from a reference protein coding sequence; and (b) selecting a DNA-protein fusion having an altered function, thereby selecting the protein having the altered function or its encoding DNA.

In preferred embodiments, the selection step involves either binding of the desired protein to an immobilized binding partner or assaying for a functional activity of the desired protein. In addition, the method may further involve repeating steps (a) and (b).

In a final aspect, the invention features a solid support including an array of immobilized molecules, each including a covalently-bonded DNA-protein fusion of the invention. In a preferred embodiment, the solid support is a microchip.

As used herein, by a "population" is meant $10^5$ or more molecules



(for example, DNA-protein fusion molecules). Because the methods of the invention facilitate selections which begin, if desired, with large numbers of candidate molecules, a "population" according to the invention preferably means more than $10^7$ molecules, more preferably, more than $10^9$, $10^{13}$, or $10^{14}$ molecules, and, most preferably, more than $10^{15}$ molecules.

By "selecting" is meant substantially partitioning a molecule from other molecules in a population. As used herein, a "selecting" step provides at least a 2-fold, preferably, a 30-fold, more preferably, a 100-fold, and, most preferably, a 1000-fold enrichment of a desired molecule relative to undesired molecules in a population following the selection step. A selection step may be repeated any number of times, and different types of selection steps may be combined in a given approach.

By a "protein" is meant any two or more naturally occurring or modified amino acids joined by one or more peptide bonds. "Protein" and "peptide" are used interchangeably herein.

By "RNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified ribonucleotides. One example of a modified RNA included within this term is phosphorothioate RNA.

By "DNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified deoxyribonucleotides.

By a "nucleic acid" is meant any two or more covalently bonded nucleotides or nucleotide analogs or derivatives. As used herein, this term includes, without limitation, DNA, RNA, and PNA.

By a "peptide acceptor" is meant any molecule capable of being added to the C-terminus of a growing protein chain by the catalytic activity of the ribosomal peptidyl transferase function. Typically, such molecules contain (i) a nucleotide or nucleotide-like moiety (for example, adenosine or an adenosine analog (di-methylation at the N-6 amino position is acceptable)), (ii)

an amino acid or amino acid-like moiety (for example, any of the 20 D- or L-amino acids or any amino acid analog thereof (for example, O-methyl tyrosine or any of the analogs described by Ellman et al., Meth. Enzymol. **202**:301, 1991), and (iii) a linkage between the two (for example, an ester, amide, or ketone linkage at the 3' position or, less preferably, the 2' position); preferably, this linkage does not significantly perturb the pucker of the ring from the natural ribonucleotide conformation. Peptide acceptors may also possess a nucleophile, which may be, without limitation, an amino group, a hydroxyl group, or a sulfhydryl group. In addition, peptide acceptors may be composed of nucleotide mimetics, amino acid mimetics, or mimetics of the combined nucleotide-amino acid structure.

By an "altered function" is meant any qualitative or quantitative change in the function of a molecule.

By "binding partner," as used herein, is meant any molecule which has a specific, covalent or non-covalent affinity for a portion of a desired DNA-protein fusion. Examples of binding partners include, without limitation, members of antigen/antibody pairs, protein/inhibitor pairs, receptor/ligand pairs (for example cell surface receptor/ligand pairs, such as hormone receptor/peptide hormone pairs), enzyme/substrate pairs (for example, kinase/substrate pairs), lectin/carbohydrate pairs, oligomeric or heterooligomeric protein aggregates, DNA binding protein/DNA binding site pairs, RNA/protein pairs, and nucleic acid duplexes, heteroduplexes, or ligated strands, as well as any molecule which is capable of forming one or more covalent or non-covalent bonds (for example, disulfide bonds) with any portion of a DNA-protein fusion.

By a "solid support" is meant, without limitation, any column (or column material), bead, test tube, microtiter dish, solid particle (for example, agarose or sepharose), microchip (for example, silicon, silicon-glass, or gold



chip), or membrane (for example, the membrane of a liposome or vesicle) to which an affinity complex may be bound, either directly or indirectly (for example, through other binding partner intermediates such as other antibodies or Protein A), or in which an affinity complex may be embedded (for example, through a receptor or channel).

The present invention provides methods for the creation of fusions between proteins and their encoding cDNAs. These constructs possess greatly enhanced chemical stability, first, due to the DNA component of the fusion and, second, due to the covalent bond linking of the DNA and protein moieties. These properties allow for easier handling of the fusion products and thereby allow selection and recognition experiments to be carried out under a range of reaction conditions. In addition, the present invention facilitates applications where a single-stranded nucleic acid portion is mandatory, for example, in hybridization assays in which the coding fusions are immobilized to a solid support. In addition, incubations may be performed under more rigorous conditions, involving high pH, elevated concentrations of multivalent metal ions, prolonged heat treatment, and exposure to various biological materials. Finally, single-stranded DNA is relatively resistant to secondary structure formation, providing a great advantage for techniques involving or requiring nucleic acid hybridization steps.

In addition, the methods of the present invention allow for the production of fusions involving DNA and protein components of any length, as well as fusion libraries of high complexity.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

## Brief Description of the Drawings

FIGURE 1 is a schematic illustration of a method for the generation

of DNA-protein fusions (Type A1) that involves ligation of a puromycin-modified DNA hairpin-like structure to an mRNA molecule.

FIGURE 2 is a schematic illustration of a method for the generation of branched hairpin structures.

FIGURE 3 is a schematic illustration of a method for the synthesis of puromycin-5'-phosphoramidite.

FIGURE 4 is a schematic illustration of a method for the generation of branched hairpin structures.

FIGURE 5 is a schematic illustration of a method for the generation of DNA-protein fusions that involves photocrosslinking of a 5'-psoralen-modified primer DNA to a suitable linker that bears a 3'-puromycin.

FIGURE 6 is a schematic illustration of exemplary methods for the chemical ligation of mRNA and DNA molecules.

FIGURE 7 is a schematic illustration of a method for the synthesis of hydrazide phosphoramidite.

FIGURE 8 is a schematic illustration of a method for the synthesis of hydrazine phosphoramidite.

FIGURE 9 is a schematic illustration of a method for the generation of DNA-protein fusions that involves chemical crosslinking of a puromycin-modified linker to the 3'-end of an mRNA molecule.

FIGURE 10 is a schematic illustration of a method for the generation of DNA-protein fusions that involves psoralen-mediated photocrosslinking of a combined linker/reverse transcription primer construct to the 3'-end of an mRNA molecule.

FIGURE 11 is a schematic illustration of an alternative method for the generation of DNA-protein fusions that involves psoralen photocrosslinking of a combined linker/reverse transcription primer construct.

FIGURE 12 is a schematic illustration of a method for the generation



of DNA-protein fusions that involves crosslinking of a reverse transcription primer to a preexisting mRNA-linker construct.

FIGURE 13 is a schematic illustration of a method for the generation of DNA-protein fusions that involves crosslinking of a reverse transcription primer to a preexisting mRNA-protein fusion.

FIGURE 14 is a schematic illustration of the oligonucleotide constructs (SEQ ID NOS: 1-6) used for the preparation of the exemplary DNA-protein fusions described herein.

FIGURE 15 is a schematic illustration of the preparation of Type C2 DNA-protein fusions.

FIGURE 16 is a photograph illustrating a product analysis of the Type C2 DNA-protein fusions.

FIGURE 17 is a schematic illustration of the preparation of Type B3 DNA-protein fusions.

FIGURE 18 is a schematic illustration of the preparation of Type B2 DNA-protein fusions.

FIGURE 19 is a photograph illustrating the resistance analysis of Type B3 DNA-protein fusions against nuclease and base treatment.

FIGURE 20 is a graph illustrating the experimentally determined half-lives of RNA- and DNA-protein fusion products in the presence of cell membrane preparations.

## Detailed Description

There are now provided below a number of exemplary techniques for the production of DNA-protein fusions, and descriptions for their use. These examples are provided for the purpose of illustrating, and not limiting, the invention.

Type A1: Template-Directed Ligation of a Puromycin-Modified Hairpin-Like Structure to an mRNA

According to a first exemplary approach, DNA-protein fusions are generated by ligating a puromycin-modified DNA hairpin-like structure to an mRNA molecule, as illustrated in Figure 1. The first step of this procedure is the attachment of puromycin to the hairpin, and this may be accomplished by a number of techniques, one of which is shown in Figure 2. By this approach, a DNA hairpin is synthesized with a puromycin-terminated side chain branching out from the DNA molecule. This construct may be generated using an asymmetric branched phosphoramidite (Clontech, Palo Alto, CA) in any standard automated DNA synthesis of the hairpin structure (see, for example, User Guide to Expedite Nucleic Acid Synthesis System, Perseptive Biosystems, Framingham, MA), followed by the addition of a 5'-phosphate using a chemical phosphorylation reagent (Glen Research, Sterling VA).

Subsequently the protecting group is selectively removed from the branch (Product Protocol for Asymmetric Branching Phosphoramidite, Clontech, Palo Alto, CA), followed by the attachment of the linker portion through standard automated DNA synthesis. Before reaching the end of the linker, the strand orientation is reversed by the addition of a few 5'-phosphoramidites (Glen Research, Sterling, VA). Finally, the synthesis is terminated through attachment of the puromycin-5'-phosphoramidite, preferably using the synthetic technique shown in Figure 3. In Figure 3, steps (a)-(c) may be carried out as described in Greene & Wuts (Protective Groups in Organic Synthesis, 2nd ed. (1991) John Wiley & Sons, Inc., New York, New York), and step (d) may be carried out as described in Beaucage (Methods in Molecular Biology, vol. 20, Protocols for Oligonucleotides and Analogs, ed. S. Agarwal (1993) Humana Press, Totowa, N.J., pp. 33-61).

Alternatively, the puromycin-modified branched hairpin may be



synthesized as shown in Figure 4. By this technique, synthesis is initiated from a puromycin-CPG solid support (Glen Research, Sterling, VA) by first synthesizing the linker portion, followed by incorporation of the branched amidite (Clontech, Palo Alto, CA) and addition of the 5'-portion of the hairpin. After deprotection of the branch, the 3'-arm of the hairpin is added by using nucleoside-5'-
phosphoramidites (Glen Research, Sterling, VA).

By either of the above approaches, in the next step, the mRNA is ligated to the hairpin, for example, using T4 DNA ligase and the 3'-overhang as a template (Sambrook, Fritsch & Maniatis Molecular Cloning (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). Ribosomal translation of the RNA then leads to protein synthesis with subsequent fusion formation (see, for example, Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302). In one particular embodiment, the branching point is located in the loop region of the hairpin. Other positions of the branching point (e.g., within the stem structure) may also be utilized. In addition, while a $dA_n$ linker of between approximately 10-60 nucleotides, and more preferably approximately 30 nucleotides, is utilized, both the length and the chemical composition (e.g., PEG (Glen Research, Sterling, VA) rather than $dA_n$) of the linker may be optimized.

In a final step, the RNA portion of the construct is reverse transcribed into cDNA (for example, as described in Sambrook, Fritsch & Maniatis, Molecular Cloning, (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) using the hairpin 3' end as a primer. Optional digestion of the mRNA by RNase H (see, for example, Sambrook, Fritsch & Maniatis Molecular Cloning, (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) yields a single stranded DNA-protein fusion.

This method also facilitates the formation of truncated DNA transcripts by adding didesoxynucleoside triphosphates during transcription (see, for example, Sanger, Science (1981) vol. 214, p. 1205-1210). Such truncated DNA-protein fusions are useful in protein display experiments (Kumelis et al., U.S.S.N. 60/080,686, filed April 3, 1998), for example, where only the 3'-region of the original message (now the 5'-region of the DNA transcript) is used for hybridization with immobilized oligonucleotide probes.

Type A2: Crosslinking of a Puromycin-Modified Linker to a Primer DNA

As an alternative to the hairpin-like construct described above, a closely related structure may also be prepared through photocrosslinking of a 5'-psoralen-modified primer DNA with a suitable linker that bears a 3'-puromycin. An exemplary crosslinking method is illustrated in Figure 5. In this method, the puromycin-bearing linker may be constructed as described, for example, in Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. The psoralen-modified primer may be generated and the photocrosslinking step carried out as described, for example, in Pieles & Englisch, Nucl. Acids Res. (1989) vol. 17, p. 285-299. The remaining steps may be carried out as described above. This approach does not require the use of non-standard nucleoside/ puromycin-5'-phosphoramidites (i.e., which were used during the automated synthesis of the hairpin-linker structure), providing an advantage over the hairpin method. Again, as above, while a $dA_n$ linker of between approximately 10-60 nucleotides, and more preferably approximately 30 nucleotides, is utilized, both the length and the chemical composition (e.g., PEG (Glen Research, Sterling, VA) rather than $dA_n$) of the linker may be optimized.

In addition, for each of the Type A1 and Type A2 methods, the



ligation reaction between the mRNA and the DNA portion of the construct may be carried out by several alternative techniques. For example, in addition to the enzymatic ligation with T4 DNA ligase described above, this step may be accomplished using chemical methods. In one particular example, the 5'-end of the hairpin may be modified with one (or multiple) amino-groups using the appropriate phosphoramidite (Clontech, Palo Alto, CA). After periodate oxidation of the 3'-end of the RNA, the two substrates may be joined through a reductive amination reaction. This is illustrated as scheme "A" in Figure 6 and is described, for example, in Lemaitre et al., Proc. Natl. Acad. Sci. USA (1987) vol. 84, p. 648-652. Alternatively, this chemical ligation step may involve carbohydrazide or hydrazine modified structures for hydrazone formation or reductive amination. These approaches are illustrated in Figure 6, respectively, as schemes "B" and "C" and are described, respectively, in Gosh et al. (Anal. Biochem. (1989) vol. 178, p. 43-51) and Proudnikov & Mirzabekov (Nucl. Acids Res. (1996) vol. 24 p. 4535-4542). Hydrazide phosphoramidite synthesis may be carried out as shown in Figure 7, and hydrazine phosphoramidite synthesis as shown in Figure 8 and as described in Greene & Wuts (Protective Groups in Organic Synthesis, 2nd ed. (1991) John Wiley & Sons, Inc., New York, New York (steps (a) and (c)), Proudnikov & Mirzabekov (Nucl. Acids Res. (1996) vol. 24 p. 4535-4542 (step b)), and Beaucage (Methods in Molecular Biology, vol. 20, Protocols for Oligonucleotides and Analogs, ed. S. Agarwal (1993) Humana Press, Totowa, N.J., pp. 33-61 (step (e)).

Types B1-B3: Chemical Crosslinking to the 3'-end of an mRNA

Yet another approach to the generation of DNA-protein fusions involves the chemical crosslinking of a puromycin-modified linker to the 3'-end of an mRNA molecule. Such crosslinking may be accomplished by a number of approaches.

One exemplary approach is shown schematically in Figure 9. In this approach ("B1"), an oligonucleotide is synthesized that bears a reactive group (e.g., one of the amino derivatives, hydrazides, or hydrazines described above) located between the primer and the linker regions. Duplex formation of the RNA and the primer site takes place immediately adjacent to this reactive group, which then is allowed to react with the periodate-oxidized 3'-end of the RNA leading to a crosslink (as shown in Figure 6 and as described above). This reaction may occur through reductive amination (Figure 6, scheme "A" or "C"; Lemaitre et al., Proc. Natl. Acad. Sci. USA (1987) vol. 84, p. 648-652; Proudnikov & Mirzabekov, Nucl. Acids Res. (1996) vol. 24 p. 4535-4542) or hydrazone formation (Figure 6, scheme "B"; Gosh et al., Anal. Biochem. (1989) vol. 178, p. 43-51). Following translation and fusion formation (Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302), the primer is extended by reverse transcriptase on the RNA template and an optional RNase H digestion step is carried out, generating the DNA-protein fusion (Figure 9).

As in methods A1 and A2 above, the strand direction of the linker portion's terminal nucleotides is reversed, which can be accomplished by the use of 5'-phosphoramidites (Glen Research, Sterling, VA) during synthesis.

In yet another exemplary crosslinking approach ("B2"), a photoreactive psoralen moiety is included in the linker as a reactive group (Figure 10). Such a construct may be synthesized using a psoralen-modified desoxynucleotide phosphoramidite (Pieles et al., Nucleic Acids Res. (1989) vol. 17, p. 8967-8978) or by incorporating a branched phosphoramidite (Clontech, Palo Alto, CA) to which a standard psoralen phosphoramidite is attached (Glen Research, Sterling, VA). Following hybridization of the linker to the target RNA, crosslink formation is achieved through irradiation with UV-light, for



example, as described in Pieles and Englisch (Nucl. Acids Res. (1989) vol. 17, p. 285-299). The resulting construct is then subjected to translation and fusion formation (Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO 98/31700; Roberts and Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302). Reverse transcription and RNase H digestion yields the final DNA-protein fusions.

Alternatively, crosslinking may be accomplished using a combined linker/reverse transcriptase primer construct as depicted in Figure 11 ("B3"). In a variant of the above approach, the psoralen moiety is not directly attached between the linker and primer region, but rather connected to a short DNA branch. This DNA portion also hybridizes to the target RNA and thus provides an optimized double-stranded environment for the psoralen to react (Pieles and Englisch, Nucl. Acids. Res. (1989) vol. 17, p. 285-299). Preparation of DNA-protein fusions using this psoralen construct may be carried out as described above.

Types C1 and C2: Crosslinking of the Reverse Transcription Primer to Preexisting mRNA-Linker Constructs

Another method for generating DNA-protein fusions is shown schematically in Figure 12. By this approach, RNA is initially ligated to a linker molecule as previously described (Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302). In a subsequent step, a suitable primer bearing a 5'-photocrosslinking reagent (e.g., psoralen, Glen Research, Sterling, VA) is annealed to the RNA-linker product. Irradiation with light furnishes a covalent crosslink between the two oligonucleotide strands (as described, for example, in Pieles & Englisch, Nucl. Acids Res. (1989) vol. 17, p. 285-299). As in methods Type A1, A2, and B1-

B3 above, translation and fusion formation may be carried out, followed by a reverse transcription step and an optional RNase H digestion step to yield DNA-protein fusions (Figure 12).

Alternatively, as shown in Figure 13, the initial steps of the above procedure may be carried out in the opposite order. This approach allows translation and fusion formation to be performed prior to crosslinking and reverse transcription. Accordingly, this method allows for the use of previously described and well established reaction conditions and components for translation and RNA-protein fusion formation.

Experimental Results

Exemplary techniques described above were carried out to demonstrate DNA-protein fusion formation. These experiments made use of the oligonucleotides depicted in Figure 14.

Model RNA substrates **1**: GGG ACA AUU ACU AUU UAC AAU UAC AAU GGA CUA CAA GGA CGA UGA CGA UAA GGG CGG CUG GUC CCA CCC CCA GUU CGA GAA GGC AUC CGC U (SEQ ID NO: 7); **2**: GGG ACA AUU ACU AUU UAC AAU UAC AAU GGA CUA CAA GGA CGA UGA CGA UAA GGG CGG CUG GUC CCA CCC CCA GUU CGA GAA GGC AUC CGC UCU UUC ACU AUA (SEQ ID NO: 8); and **3**: GGG ACA AUU ACU AUU UAC AAU UAC AAU GGA CUA CAA GGA CGA UGA CGA UAA GGG CGG CUG GUC CCA CCC CCA GUU CGA GAA GGC AUC CGC UAU UUA AAA AAA AAA AAA AAA AAA A (SEQ ID NO: 9) were synthesized by T7 transcription (Megashortscript transkiption kit, Ambion, Austin, TX) using appropriate dsDNA templates. Following transcription, the RNAs were purified by denaturing polyacrylamide gel electrophoresis.

The modified oligonucleotides **4**: 5' pd(AAA AAA AAA ACG GCT



ATA TAA AAA AAA CC)- Pu (SEQ ID NO: 10); **5**: 5' psoralen C2-TAG CCG TTT TTT TTT TAG CGG ATG C (SEQ ID NO: 11); **6**: 5' d(cgt agg cga gaa agt gat)-branch[psoralen C6]-d(AAA AAA AAA AAA AAA AAA AAA AAA AAA CC)-Pu (SEQ ID NO: 12); and **7**: 5' ggt caa gct ctt-branch[5' psoralen C6-TAG CGG ATG C 3'] $spacer_6$ CC-Pu (SEQ ID NO: 13) [[uppercase = standard DNA-3'-phosphoramidites; lowercase = DNA-5'-phosphoramidites; spacer = spacer-9 phosphoramidite; Pu = puromycin-CPG (all from Glen Research, Sterling, VA); branch = asymmetric branching amidite (Clontech, Palo Alto, CA)] were synthesized on an Expedite Synthesizer Model 8909 (PerSeptive Biosystems, Framingham, MA) according to recommended protocols for the corresponding phosphoramidites. For the branched constructs **6** and **7**, the main chain was synthesized first and concluded with a final capping step. Next, the levulinyl protecting group was removed from the branching unit through treatment with 0.5 M hydrazine monohydrate in pyridine-acetic acid for 15 minutes at room temperature. Automated synthesis was then resumed and the side chain sequences (indicated in square brackets) were attached. The oligos were fully deprotected in concentrated ammonium hydroxide for 8 hours at 55°C and purified by denaturing gel electrophoresis.

The DNA sequences **8**: d(TTT TTT TTT TAG CGG ATG C) (SEQ ID NO: 14) and **9**: d(TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT) (SEQ ID NO: 15) were purchased from Oligos etc. (Wilsonville, OR) and used without further purification.

Type C2 DNA-Protein Fusion Formation

Type C2 DNA-protein fusion formation was demonstrated as follows (Figure 15). RNA **1** and linker **4** were hybridized to template DNA **8** and enzymatically ligated by T4 DNA ligase as previously described (Szostak et al.,

U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts and Szostak, Proc. Natl. Acad. Sci. USA (1997) Vol. 94, p. 12297-12302). After purification by electrophoresis on a denaturing polyacrylamide gel, the resulting mRNA-linker construct was used as a template for *in vitro* translation using rabbit reticulocyte lysate kits from Ambion. Reaction mixtures contained 50 pmole ligated mRNA **10**, 10 mM creatine phosphate, 150 mM potassium acetate, 0.5 mM magnesium chloride, 0.1 mM of each amino acid exept methionine, 150 µCi [$^{35}$S] methionine (Amersham, Arlington Heights, IL) and 67% v/v of lysate in a total volume of 300 µl and were carried out for 30 minutes at 30°C. To promote the subsequent fusion formation, KCl and $MgCl_2$ were added to 590 mM and 50 mM final concentrations, respectively, in a volume of 500 µl. Incubation was continued for 60 minutes at 20°C. Products were isolated by diluting the lysate into 10 ml of binding buffer (100 mM Tris pH 8.0, 10 mM EDTA, 1 M NaCl, 0.25% v/v Triton X-100) and adding 10 mg oligo-dT cellulose type 7 (Pharmacia, Piscataway, NJ). Samples were rotated for 60 minutes at 4°C, and the solid support was then washed with 5 ml ice-cold binding buffer that was devoid of EDTA, followed by elution with 100 µl aliquots of water. Fusion product was found in fractions 2 and 3, and these fractions were combined. The total yield of fusion **11** was determined by scintillation counting of the incorporated [$^{35}$S] methionine to be 1.6 pmole (3.2% of input RNA).

For the conversion of the RNA-protein fusions **11** into DNA-protein fusions **13**, the following reactions were performed (Figure 15). First, 20 µl of the above oligo-dT-purified material **11** was mixed with 0.5 µl primers **5** (50 µM) and 6 µl first strand buffer (Superscript II kit from GibcoBRL; 250 mM Tris-HCl pH 8.3, 375 KCl, 15mM $MgCl_2$) and briefly heated to 80°C for 2 minutes, followed by slowly cooling to 0°C. Psoralen photocrosslink formation was induced by irradiating the sample for 15 minutes at 0°C with $\lambda >$ 310 nm [450 W medium pressure immersion lamp (ACE Glass, Vineland, NJ) equipped with a Pyrex absorption sleeve in a Quartz immersion well]. Next, 0.6 µl of a dNTP mix (25 mM each), 3 µl of 0.1M DTT, and 0.4 µl (80 units) Superscript II reverse transcriptase were added, and cDNA synthesis was carried out for 60 minutes at 42°C. The RNA portion was then removed by continuing incubation for 60 minutes at 37°C after addition of 0.5 µl (1 unit) RNase H (Promega, Madison, WI). Finally, double-stranded DNA **14** was generated by adding 50 pmoles of primer **9** and incubating for another 60 minutes at 42°C. Control reactions with non-crosslinked samples were performed as indicated in Figure 15. Product analysis was performed by electrophoresis on denaturing 6% TBE-Urea gels (Novex, San Diego, CA), followed by visualization of the [$^{35}$S]-labelled product bands by exposure on a phosphorimager screen (Figure 16).



Samples were applied to the gel in the same order as they appear in Figure 15, beginning with RNA-protein fusion **11** and following the reaction pathway with and without having been photocrosslinked. As indicated in Figure 16, the gel mobilities correspond well with the expected behavior and clearly confirm the constitution of DNA-protein fusion **13**.

Type B3 DNA-Protein Fusion Formation

Type B3 DNA-protein fusion formation was demonstrated as follows (Figure 17). The branched linker construct **7** (5 µM) was annealed to the target RNA **3** (2.5 µM) in 25 mM Tris buffer pH 7.0 containing 100 mM NaCl and crosslinked by irradiation for 15 minutes at room temperature in a borosilicate glass vial (Kimble/Kontes, Vineland, NJ) using a handheld multiwavelength UV lamp model UVGL-25 (UVP, Upland, CA) set to long wave. Product analysis was performed by electrophoresis on a 6% TBE-Urea polyacrylamide gel followed by visualization by UV shadowing. These results indicated nearly

quantitative conversion of the starting material (gel "A" in Fig. 17). The photoligated product RNA was used for *in vitro* translation without further separation from remaining unligated RNA and excess linker. *In vitro* translation and fusion formation reactions were performed as described for Type C2 above, with 100 pmole input RNA in a 300 µl total volume. After purification on oligo-dT cellulose, 5.5 pmole RNA-fusion **15** was obtained. Its conversion into single-stranded and double-stranded DNA-protein fusions **16** and **17**, respectively, was done by reverse transcription (Superscript II kit, GibcoBRL, Grand Island, NY) and RNase H (Promega, Madison, WI) treatment as described for Type C2 fusions (gel "B" in Fig. 17).

Type B2 DNA-Protein Fusion Formation

Type B2 DNA-protein fusion formation was demonstrated as outlined in Figure 18. Specifically, following the procedure outlined for Type B3 fusions above, RNA **2** was crosslinked to linker **6**. Following denaturing polyacrylamide electrophoresis, the ligated product **18** was isolated in 12% yield. *In vitro* translation, fusion formation, and preparation of DNA-protein fusions **19** were carried out as described for Type B3 fusions above, with similar efficiencies of fusion formation.

DNA-Protein Fusion Stability Tests

To evaluate the nuclease and base resistance of DNA fusions in comparison with the corresponding RNA fusions, the following experiments were carried out. To 10 µl DNA-fusion **16** (Type B3) or RNA-fusion **15** in reverse transcription buffer was added either 0.2 µl (0.4 units) RNase H, 0.2 µl (2 units) RNase I, 0.2 µl (0.6 units) T4 DNA polymerase (3'-5' exonuclease activity), or 2.5 µl of 2.0 M NaOH. Samples were incubated for 30 minutes at 37°C and then analyzed on a 4-12% NuPage gel (Novex, San Diego, CA)



followed by autoradiography. Results are shown in Figure 19 and confirm the increased stability of DNA fusions against ribonucleases and base treatment.

To test stability of DNA fusion constructs in biological media, 5 nM of either RNA fusions **11** or **12**, or DNA fusions **13** or **14** (Type C2) were incubated with 3 µg/µl CHO-K1 cell membranes (Receptor Biology, Beltsville, MD) in 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, and 10 mM DTT at room temperature. Additional samples of RNA fusions **11** and **12** were prepared containing 20 mM vanadyl ribonucleside complex ("VRC") to inhibit ribonuclease activity. Aliquots were taken after 0, 5, 15, 30, 60, 120 minutes, and 24 hours and analyzed by electrophoresis on 4-12% NuPage gels (Novex) followed by exposure on a phosphorimager screen. The relative amounts of remaining fusion were plotted against incubation time and half-lives graphically extracted from the resulting curves. As indicated in Figure 20, all constructs showed more than 50% decay during the initial two hour period except for dsDNA fusion **14**, which appeared to be entirely stable under the conditions tested. Following a 24 hour incubation, all fusion constructs were completely degraded due to either nuclease or protease activity.

In Vitro Selection of Desired Proteins

The DNA-protein fusions described herein may be used in any selection method for desired proteins, including molecular evolution and recognition approaches. Exemplary selection methods are described, for example, in Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302; Lipovsek et al., U.S.S.N. 60/096,818 and U.S.S.N. 09/374,962; and Kuimelis et al. U.S.S.N. 60/080,686 and U.S.S.N. 09/282,734, all hereby incorporated by reference.

Use

The DNA-protein fusions described herein may be used for any application previously described or envisioned for RNA-protein fusions. Commercial uses include the isolation of polypeptides with desired properties through *in vitro* evolution techniques (see, for example, Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302)), screening of cDNA libraries that are derived from cellular mRNA (see, for example, Lipovsek et al., U.S.S.N. 60/096,818, filed August 17, 1998), and the cloning of new genes on the basis of protein-protein interactions (Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09,247,190; Szostak et al., WO98/31700), as well as the use of these fusions in protein display experiments (Kuimelis et al. U.S.S.N. 60/080,686 and U.S.S.N. 09/282,734). In addition, the DNA-protein fusions described herein may be used in binding and molecular recognition assays that involve biological materials that presumably contain ribonucleases, such as whole cells, lysates, or biological fluids. These DNA-protein fusions may be used for any appropriate therapeutic, diagnostic, or research purpose, particularly in the pharmaceutical and agricultural areas.

What is claimed is:



Claims

1. A method for generating a DNA-protein fusion, said method comprising:

(a) linking a nucleic acid primer to an RNA molecule, said primer being bound to a peptide acceptor;

(b) translating said RNA to produce a protein product, said protein product being covalently bound to said primer; and

(c) reverse transcribing said RNA to produce a DNA-protein fusion.

2. A method for generating a DNA-protein fusion, said method comprising:

(a) generating an RNA-protein fusion;

(b) hybridizing a nucleic acid primer to said fusion; and

(c) reverse transcribing said RNA to produce a DNA-protein fusion.

3. The method of claim 1 or 2, said method further comprising treating the product of step (c) to remove said RNA.

4. The method of claim 3, wherein said treating comprises contacting the product of step (c) with RNase H under conditions sufficient to digest said RNA.

5. The method of claim 1 or 2, wherein said nucleic acid primer is a DNA primer.

6. The method of claim 1 or 2, wherein said translating step is carried out in vitro.

7. The method of claim 1, wherein said peptide acceptor is puromycin.

8. The method of claim 1, wherein said nucleic acid primer has a hairpin structure.

9. The method of claim 1 or 2, wherein said primer further comprises a photocrosslinking agent.

10. The method of claim 9, wherein said photocrosslinking agent is psoralen.

11. The method of claim 9, wherein said primer is crosslinked to an oligonucleotide which is bound to a peptide acceptor.

12. The method of claim 9, wherein said primer is hybridized to said RNA molecule and said linking step is carried out by photocrosslinking.

13. A molecule comprising a DNA covalently bonded to a protein through a peptide acceptor.

14. The molecule of claim 13, wherein said peptide acceptor is puromycin.

15. The molecule of claim 13, wherein said protein comprises at least 10 amino acids.

16. A molecule comprising a DNA covalently bonded to a protein,



said protein comprising at least 10 amino acids.

17. The molecule of claim 13 or 16, wherein said protein comprises at least 30 amino acids.

18. The molecule of claim 17, wherein said protein comprises at least 100 amino acids.

19. The molecule of claim 13 or 16, wherein said protein is encoded by said DNA.

20. The molecule of claim 13 or 16, wherein said protein is entirely encoded by said DNA.

21. The molecule of claim 13 or 16, wherein said molecule further comprises a ribonucleic acid covalently bonded to said DNA.

22. The molecule of claim 21, wherein said protein is encoded by said ribonucleic acid.

23. The molecule of claim 13 or 16, wherein said DNA is double stranded.

24. A population of at least $10^5$ DNA-protein fusions, each fusion comprising a DNA covalently bonded to a protein.

25. The population of claim 24, wherein said population comprises at least $10^{14}$ fusions.

26. The population of claim 24, wherein said protein comprises at least 10 amino acids.

27. The population of claim 26, wherein said protein comprises at least 30 amino acids.

28. The population of claim 27, wherein said protein comprises at least 100 amino acids.

29. The population of claim 24, wherein said protein is encoded by said covalently bonded DNA.

30. The population of claim 24, wherein said protein is entirely encoded by said covalently bonded DNA.

31. The population of claim 24, wherein said fusions further comprise a ribonucleic acid covalently bonded to said DNA.

32. The population of claim 31, wherein said protein is encoded by said ribonucleic acid.

33. The population of claim 24, wherein said DNA is covalently bonded to said protein through a peptide acceptor.

34. The population of claim 33, wherein said peptide acceptor is puromycin.

35. The population of claim 24, wherein said DNA is double



stranded.

36. A method for the selection of a desired protein or its encoding DNA, comprising the steps of:

a) providing a population of DNA-protein fusions, each comprising a DNA covalently bonded to a candidate protein; and

b) selecting a desired DNA-protein fusion, thereby selecting said desired protein or DNA.

37. A method for the selection of a protein, or its encoding DNA, having an altered function relative to a reference protein, comprising the steps of:

a) producing a population of candidate DNA-protein fusions, each comprising a DNA covalently bonded to a candidate protein and having a candidate protein coding sequence which differs from said reference protein coding sequence; and

b) selecting a DNA-protein fusion having an altered function, thereby selecting said protein having said altered function or its encoding DNA.

38. The method of claim 36 or 37, wherein said protein is encoded by said DNA.

39. The method of claim 36 or 37, wherein said DNA-protein fusion further comprises a ribonucleic acid covalently bonded to said DNA.

40. The method of claim 39, wherein said protein is encoded by said ribonucleic acid.

41. The method of claim 36 or 37, wherein said protein comprises at least 10 amino acids.

42. The method of claim 41, wherein said protein comprises at least 30 amino acids.

43. The method of claim 42, wherein said protein comprises at least 100 amino acids.

44. The method of claim 36 or 37, wherein said DNA is covalently bonded to said protein through a peptide acceptor.

45. The method of claim 44, wherein said peptide acceptor is puromycin.

46. The method of claim 36 or 37, wherein said population of candidate DNA-protein fusions comprises at least $10^5$ different DNA molecules.

47. The method of claim 46, wherein said population of candidate DNA-protein fusions comprises at least $10^{14}$ different DNA molecules.

48. The method of claim 36 or 37, wherein said selection step comprises binding of said desired protein to an immobilized binding partner.

49. The method of claim 36 or 37, wherein said selection step comprises assaying for a functional activity of said desired protein.



50. The method of claim 36 or 37, wherein said method further comprises repeating steps (a) and (b).

51. The method of claim 36 or 37, wherein said DNA is double stranded.

52. A solid support comprising an array of immobilized molecules, each comprising a DNA covalently bonded to a protein.

53. The solid support of claim 52, wherein said protein is encoded by said DNA.

54. The solid support of claim 52, wherein said solid support is a microchip.

## DNA-Protein Fusion: Type A1

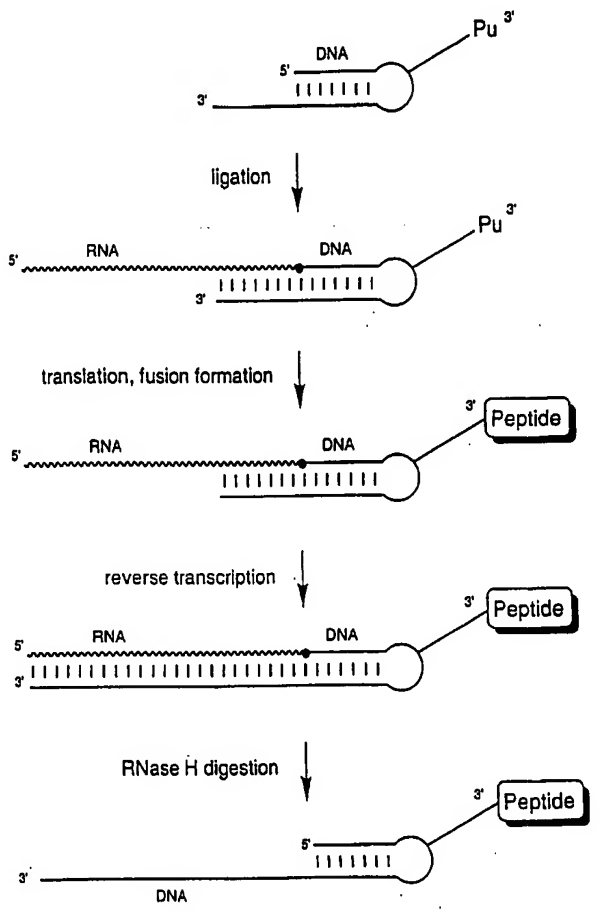



FIG. 1



## Synthesis of Branched Hairpin, Variant 1

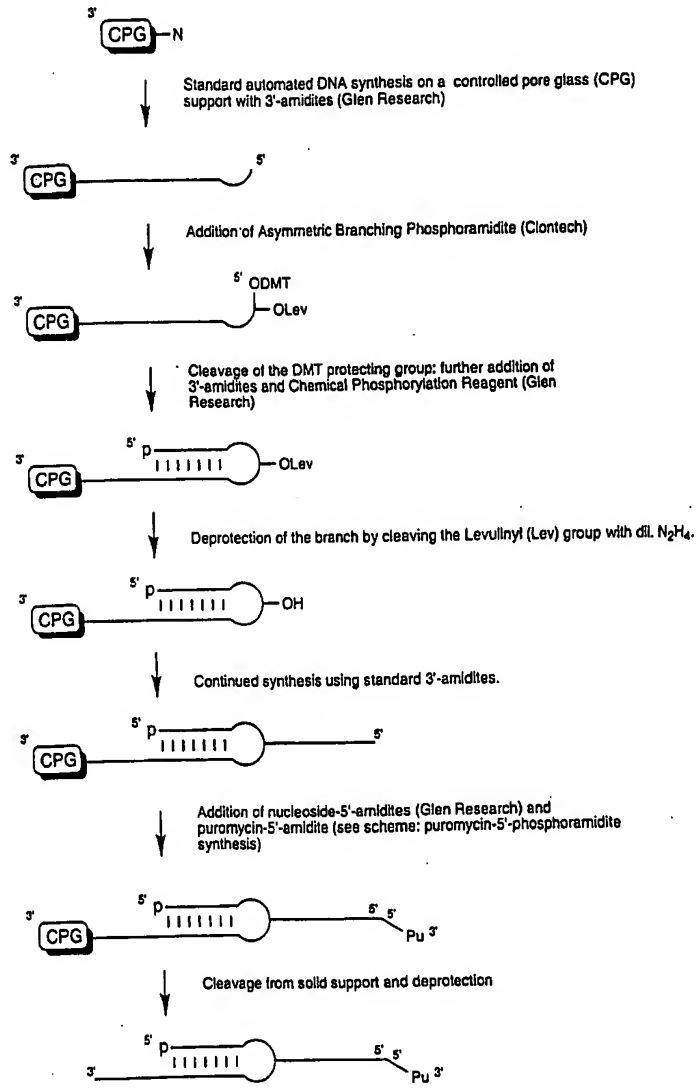



FIG. 2

## Puromycin-5'-Phosphoramidite Synthesis

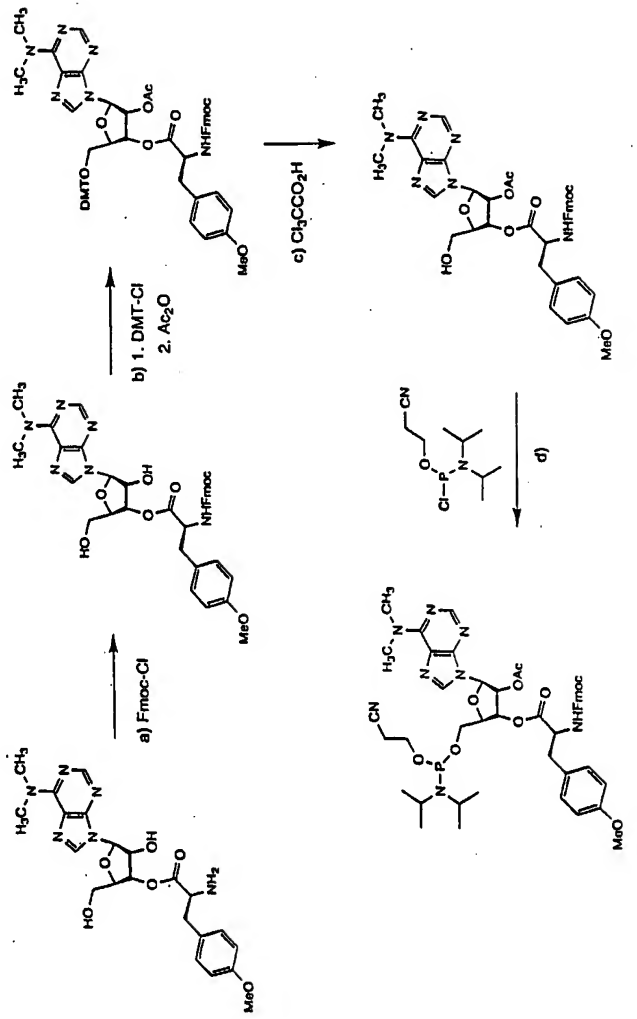



FIG. 3



## Synthesis of Branched Hairpin, Variant 2

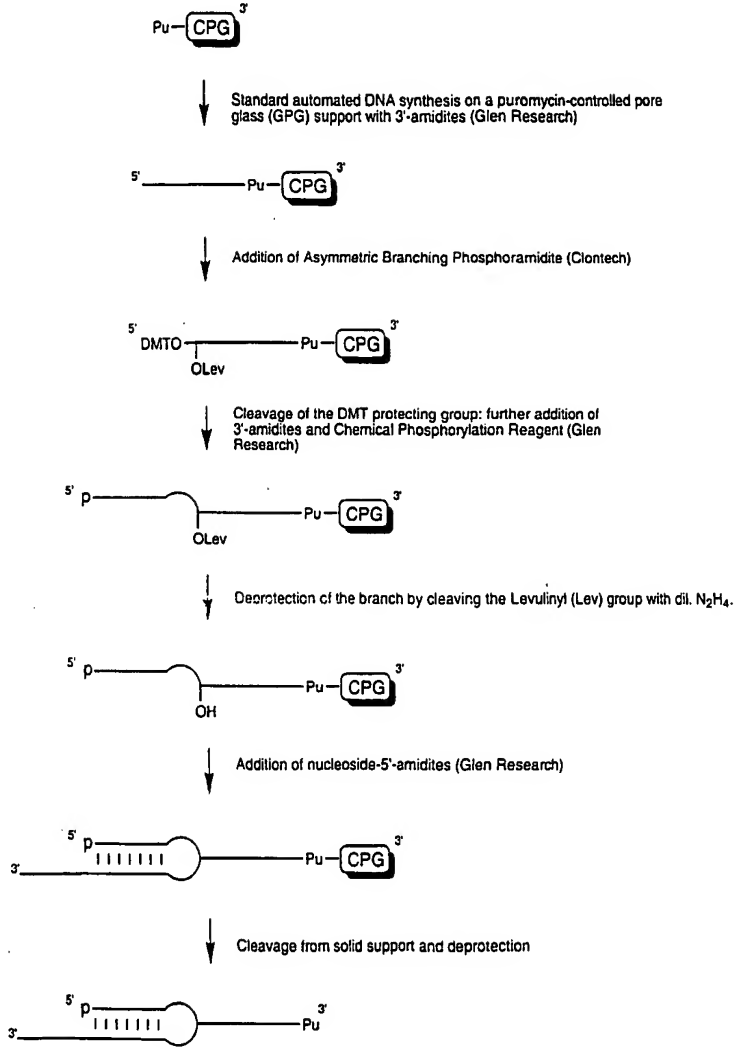



FIG. 4

## DNA-Protein Fusion: Type A2

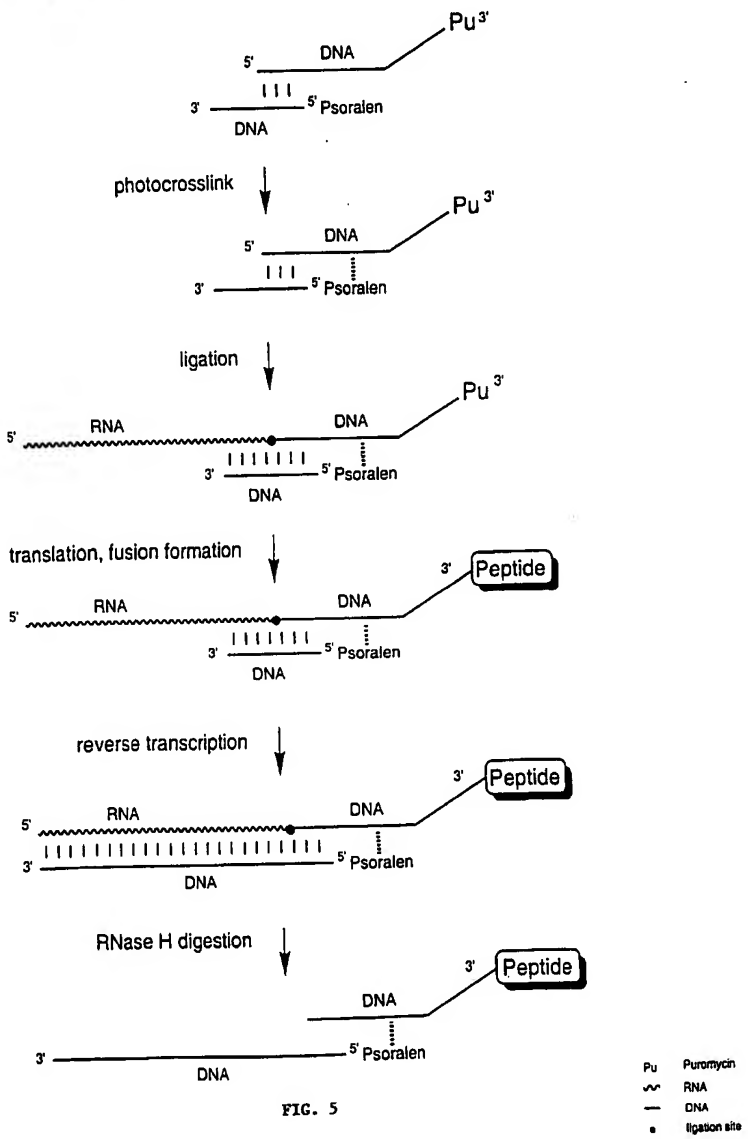



FIG. 5



## Chemical Ligation

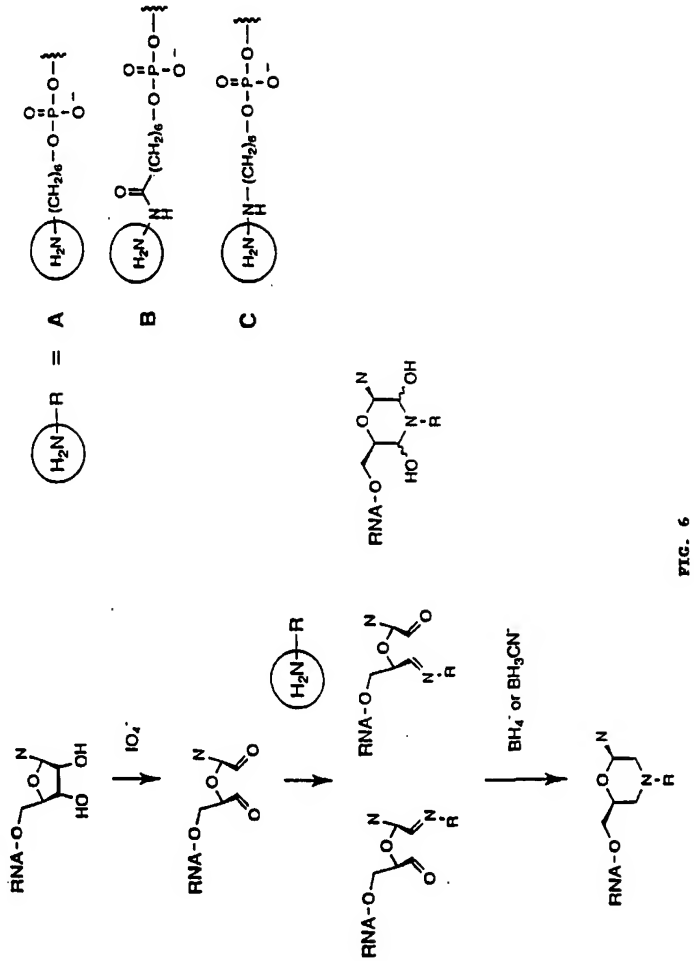



FIG. 6

## Hydrazide Phosphoramidite Synthesis

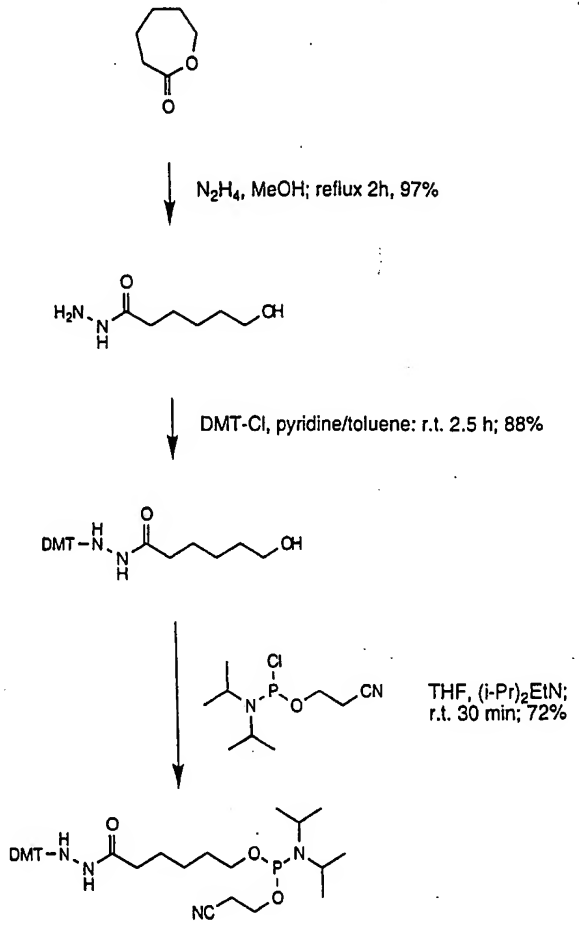



FIG. 7



## Hydrazine Phosphoramidite Synthesis

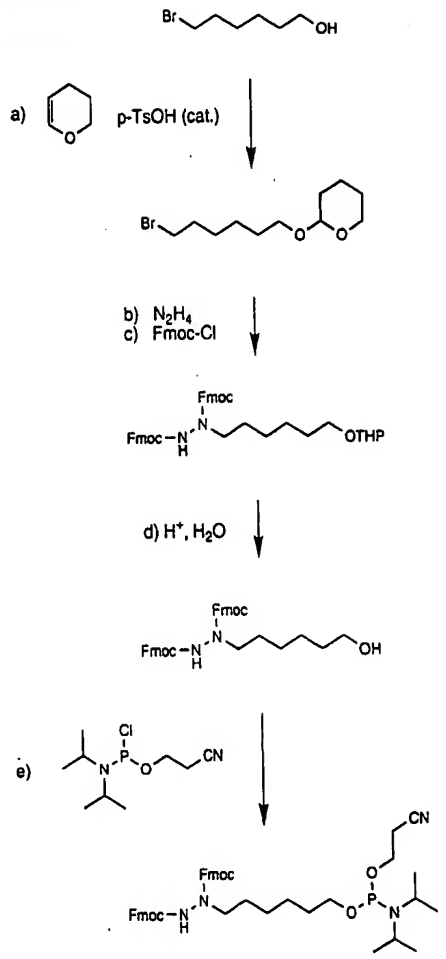



FIG. 8

## DNA-Protein Fusion: Type B1

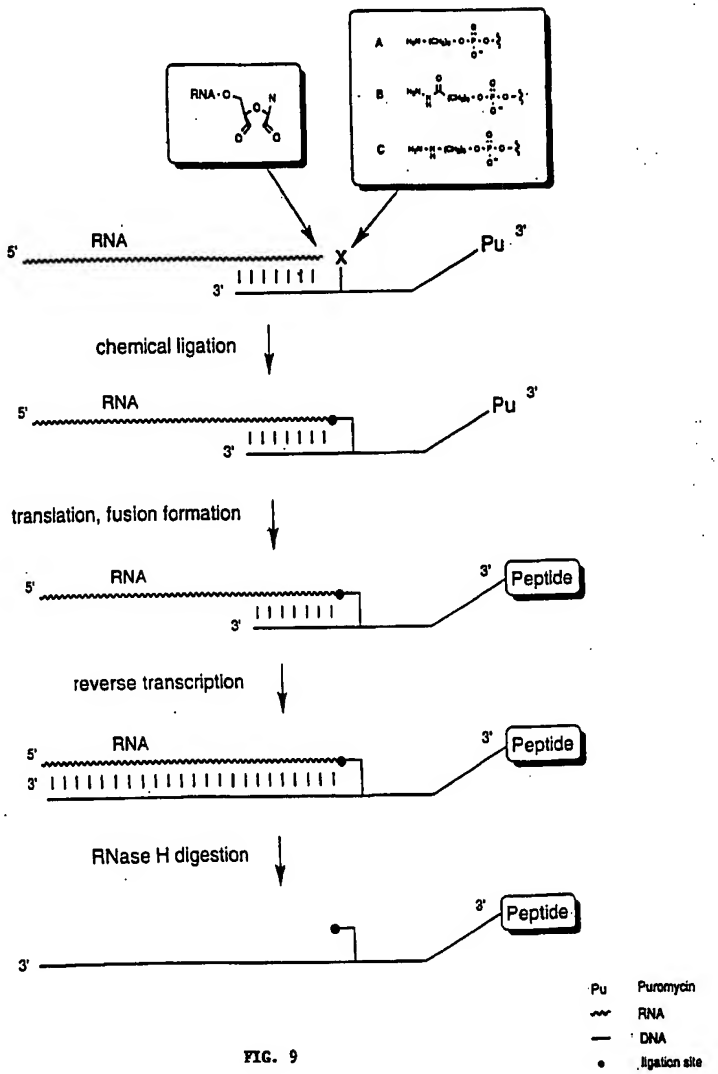



FIG. 9



## DNA-Protein Fusion: Type B2

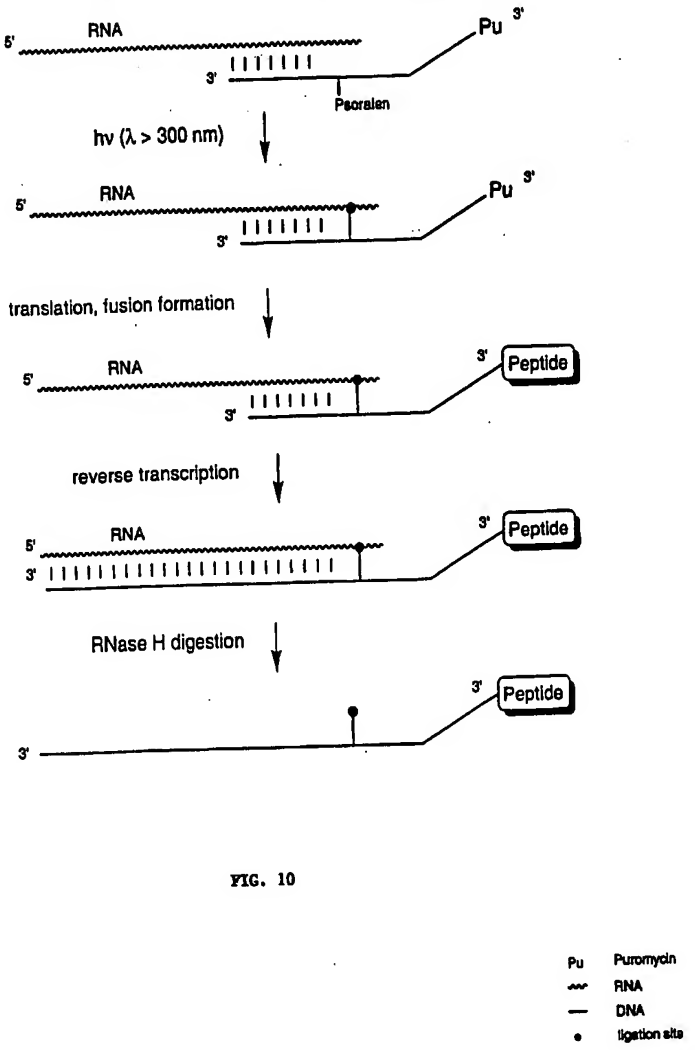



FIG. 10

## DNA-Protein Fusion: Type B3

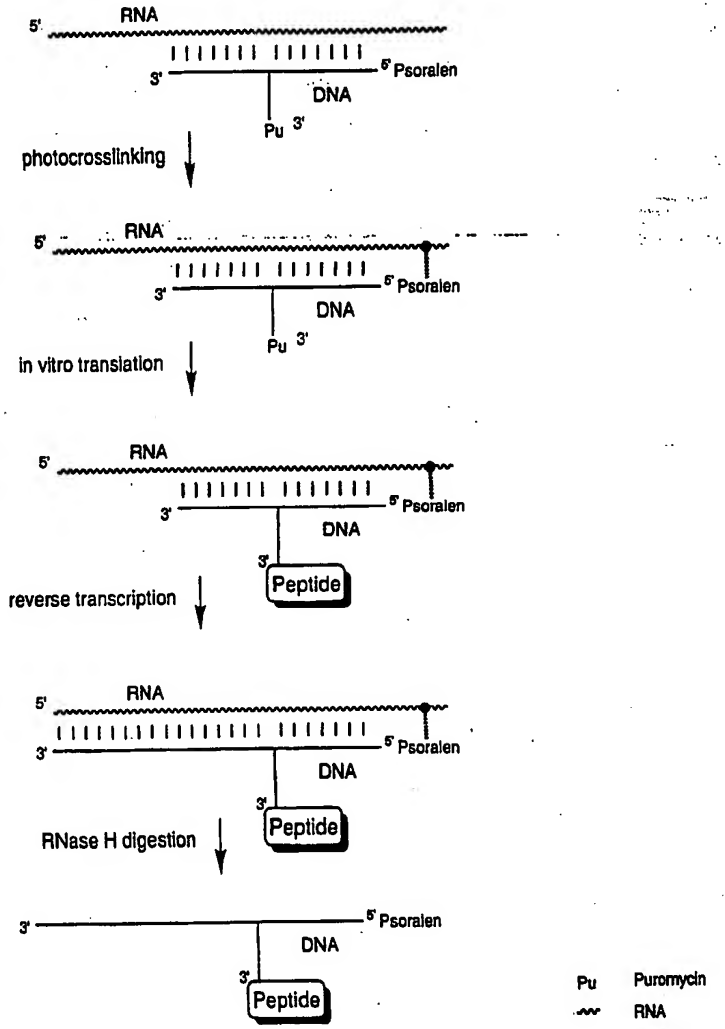



FIG. 11



## DNA-Protein Fusion: Type C1

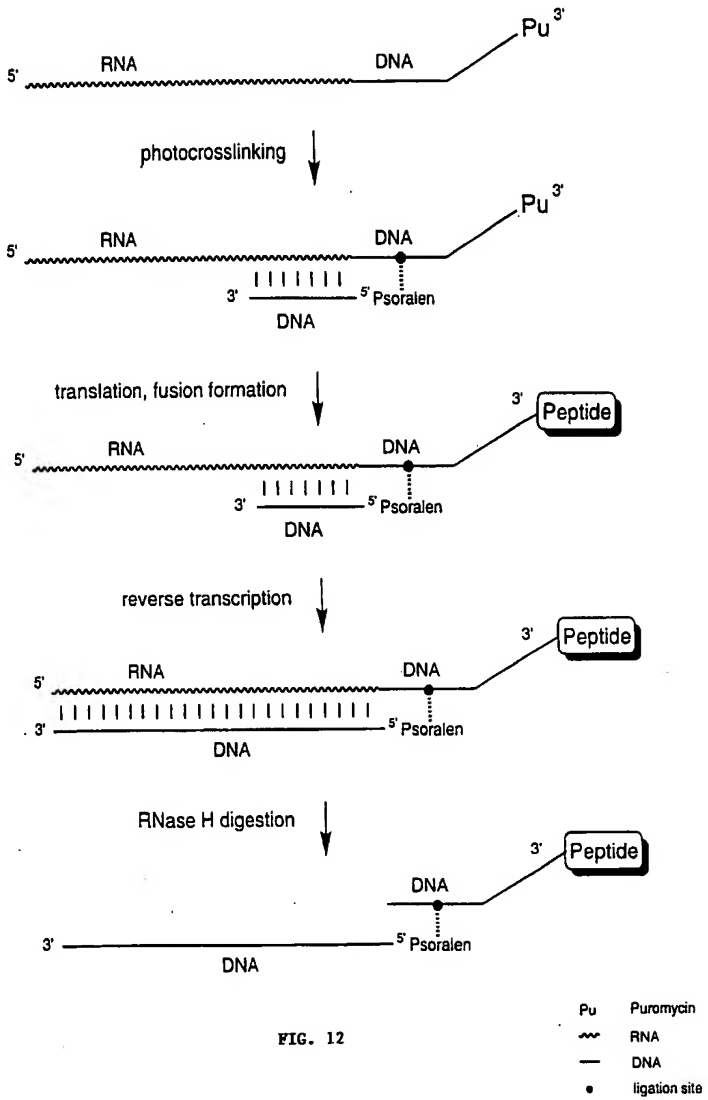



FIG. 12

## DNA-Protein Fusion: Type C2

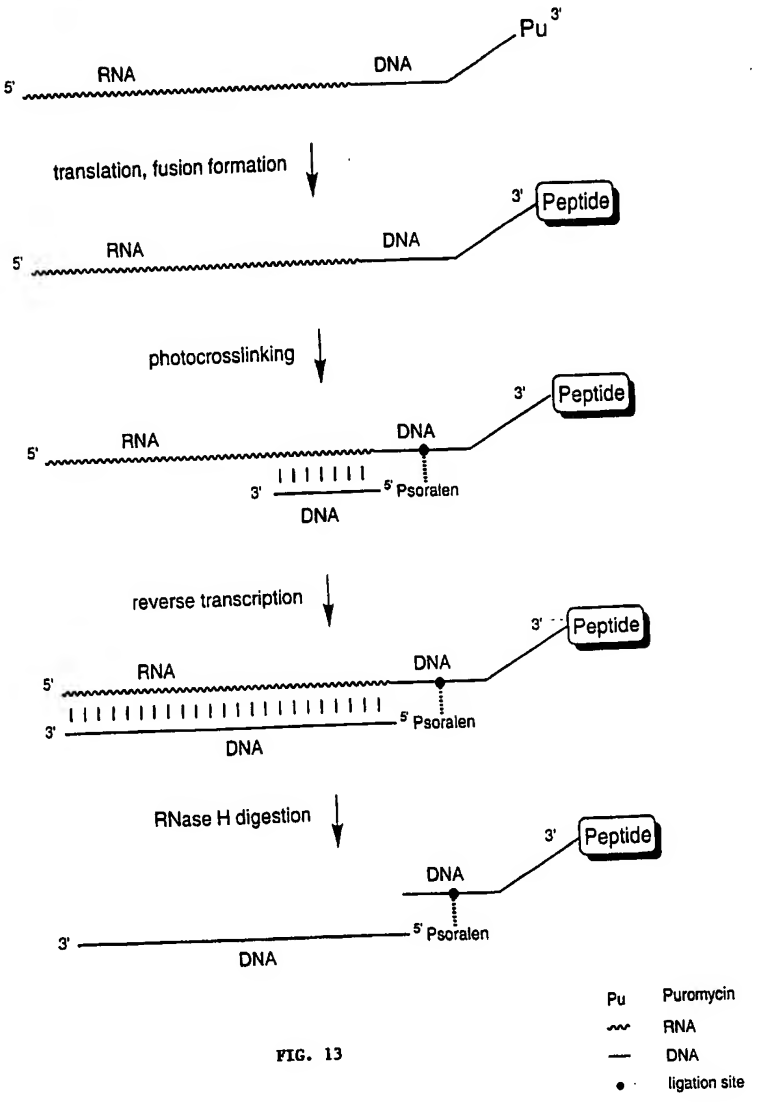
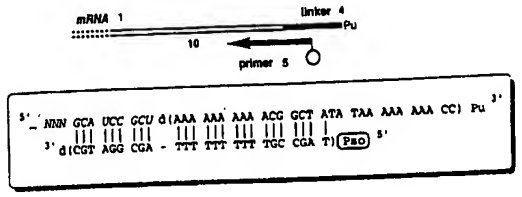


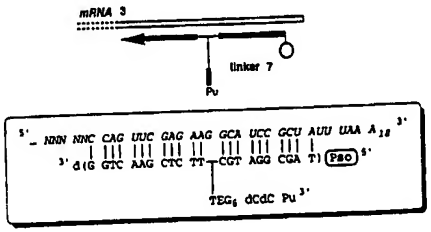

FIG. 13



Type C2




Type B3




Type B2

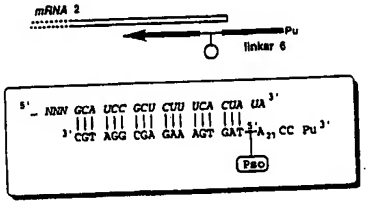



FIG. 14

Type C2

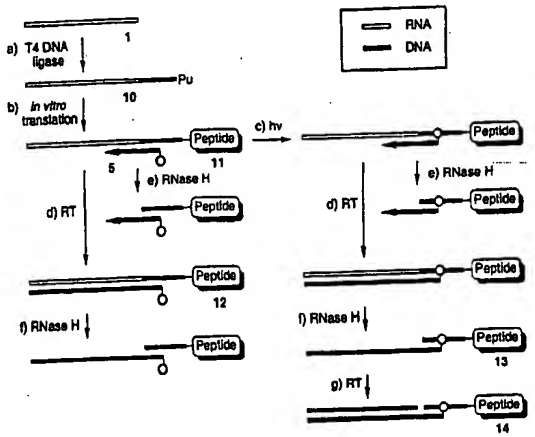



FIG. 15



Type C2

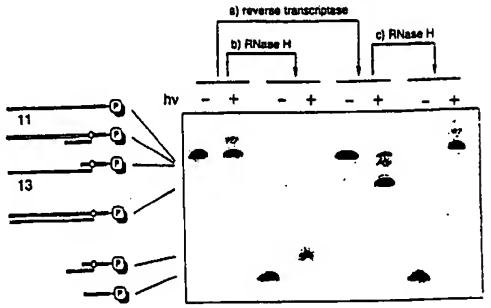



FIG. 16

## Type B3

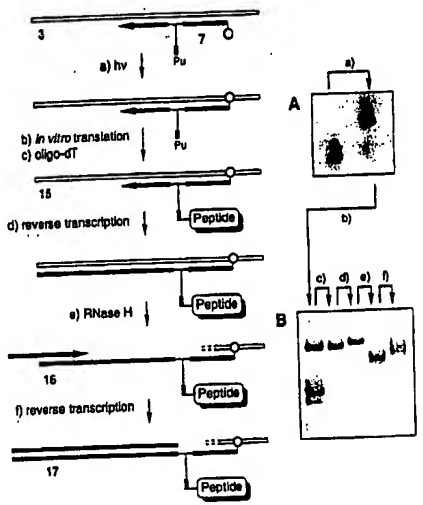



FIG. 17



## Type B2

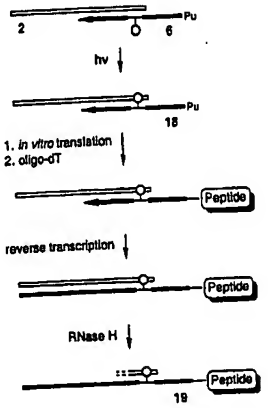



FIG. 18

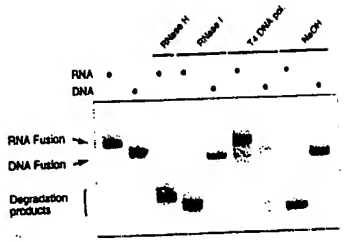



FIG. 19



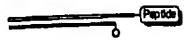
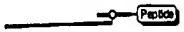
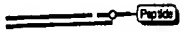
| Fusion construct | | | VRC | $t_{1/2}$ |
|---|---|---|---|---|
| ssRNA |  | 11 | - | 35 min |
| | | | + | 120 min |
| dsRNA |  | 12 | - | 12 min |
| | | | + | 50 min |
| ssDNA |  | 13 | - | 80 min |
| dsDNA |  | 14 | - | 120 min « $t_{1/2}$ < 24 h |

FIG. 20

## INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/28472

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) :C2Q 1/68; C12P 19/34; C07H 21/00, 21/02,21/04
US CL :435/6, 91.1, 91.2, 91.21; 536/22.1, 23.1, 23.4

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1, 91.2, 91.21; 536/22.1, 23.1, 23.4

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST USPATFULL, EPO,JPO,WPIDS; STN-MEDLINE,CAPLUS,EMBASE,BIOSIS, CANCERLIT
search terms: dna-protein fusion covalent,bonding

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 98/31700 A1 (THE GENREAL HOSPITAL CORPORATION) 23 JULY 1998 (23.07.98), see entire document, especially pages 2-3 | 1-54 |
| X | ROBERTS et al. RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. USA. November 1997, vol. 94, pages 12297-12302, see entire document. | 1-54 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 05 APRIL 2000 | 18 APR 2000 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer<br>JEFFREY SIEW<br>Telephone No. (703) -308-0196 |

Form PCT/ISA/210 (second sheet) (July 1998) *

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 June 2000 (08.06.2000)

PCT

(10) International Publication Number
WO 00/32823 A1

(51) International Patent Classification[7]: C12Q 1/68, C12P 19/34, C07H 21/00, 21/02, 21/04

(21) International Application Number: PCT/US99/28472

(22) International Filing Date: 2 December 1999 (02.12.1999)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/110,549 2 December 1998 (02.12.1998) US

(71) Applicant: PHYLOS, INC. [US/US]; 128 Spring Street, Lexington, MA 02421 (US).

(72) Inventors: LOHSE, Peter; 50 Golden Ball Road, Weston, MA 02493 (US). KURZ, Markus; 62 Kensington Street, West Newton, MA 02460 (US). WAGNER, Richard; 1007 Lowell Road, Concord, MA 01742 (US).

(74) Agent: ELBING, Karen, L.; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110-2214 (US).

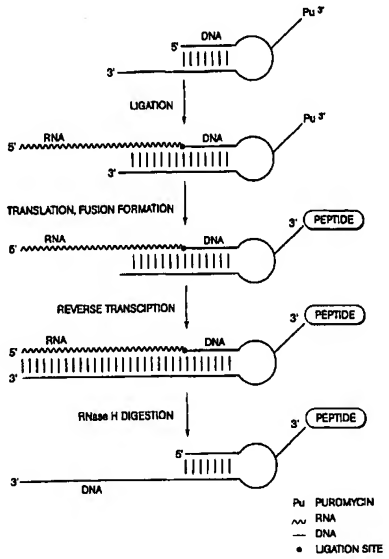
(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent

*[Continued on next page]*

(54) Title: DNA-PROTEIN FUSIONS AND USES THEREOF

(57) Abstract: The invention provides methods for covalently tagging proteins with their encoding DNA sequences as shown in the Figure. These DNA-protein fusions may be used in molecular evolution and recognition techniques.




WO 00/32823 A1

WO 00/32823 A1

(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *With international search report.*

**(48) Date of publication of this corrected version:**
29 March 2001

**(15) Information about Correction:**
see PCT Gazette No. 13/2001 of 29 March 2001, Section II

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



# DNA-PROTEIN FUSIONS AND USES THEREOF

## Background of the Invention

In general, the invention features DNA-protein fusions and their uses, particularly for the selection of desired proteins and their corresponding nucleic acid sequences.

Recently, a combinatorial method was developed for the isolation of proteins with desired properties from large pools of proteins (Szostak et al., U.S.S.N. 09/007,005; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302). By this method, the protein portion is linked to its encoding RNA by a covalent chemical bond. Due to the covalent nature of this linkage, selection experiments are not limited to the extremely mild reaction conditions that must be used for approaches that involve non-covalent complex formation such as ribosome display (Hanes & Plückthun, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 4937-4942; He & Taussig, Nucl. Acids Res. (1997) vol. 25, p 5132-5143). However, precautions do need to be taken during the selection process to minimize RNA degradation, since the accidental cleavage of ribo-bonds can result in the irreversible loss of encoded information. For this reason, these selection procedures are typically carried out using reaction media and equipment that are free of ribonucleases or other deleterious contaminants.

## Summary of the Invention

The present invention provides methods for covalently tagging proteins with their encoding DNA sequences. These DNA-protein fusions, which may be used in molecular evolution and recognition techniques, are chemically more stable than RNA-protein fusions and therefore provide a number of advantages (as discussed in more detail below).

Accordingly, in general, the invention features methods for generating DNA-protein fusions. A first method involves: (a) linking a nucleic acid primer to an RNA molecule (preferably, at or near the RNA 3' end) , the primer being bound to a peptide acceptor (for example, puromycin); (b) translating the RNA to produce a protein product, the protein product being covalently bound to the primer; and (c) reverse transcribing the RNA to produce a DNA-protein fusion.

A second method involves: (a) generating an RNA-protein fusion; (b) hybridizing a nucleic acid primer to the fusion (preferably, at or near the RNA 3' end); and (c) reverse transcribing the RNA to produce a DNA-protein fusion.

In a preferred embodiment of the above methods, the method may further involve treating the product of step (c) to remove the RNA (for example, by contacting the product of step (c) with RNase H under conditions sufficient to digest the RNA). In additional preferred embodiments, the nucleic acid primer is a DNA primer; the translating step is carried out in vitro; and the nucleic acid primer has a hairpin structure. In addition, the primer may further include a photocrosslinking agent, such as psoralen, and the primer may be crosslinked to an oligonucleotide which is bound to a peptide acceptor or, alternatively, may be hybridized to the RNA molecule, followed by a linking step that is carried out by photocrosslinking.

In related aspects, the invention also features a molecule including a DNA covalently bonded to a protein (preferably, of at least 10 amino acids) through a peptide acceptor (for example, puromycin), as well as a molecule including a DNA covalently bonded to a protein, in which the protein includes at least 10 amino acids.

In preferred embodiments of both of these aspects, the protein includes at least 30 amino acids, more preferably, at least 100 amino acids, and



may even include at least 200 or 250 amino acids. In other preferred embodiments, the protein is encoded by the DNA and is preferably entirely encoded by the DNA; the molecule further includes a ribonucleic acid covalently bonded to the DNA; the protein is encoded by the ribonucleic acid; and the DNA is double stranded.

In another related aspect, the invention features a population of at least $10^5$, and preferably, at least $10^{14}$, DNA-protein fusions of the invention, each fusion including a DNA covalently bonded to a protein.

In addition, the invention features selection methods which utilize the DNA-protein fusions described herein. A first selection method involves the steps of: (a) providing a population of DNA-protein fusions, each including a DNA covalently bonded to a candidate protein; and (b) selecting a desired DNA-protein fusion, thereby selecting the desired protein or DNA.

A second selection method involves the steps of: (a) producing a population of candidate DNA-protein fusions, each including a DNA covalently bonded to a candidate protein and having a candidate protein coding sequence which differs from a reference protein coding sequence; and (b) selecting a DNA-protein fusion having an altered function, thereby selecting the protein having the altered function or its encoding DNA.

In preferred embodiments, the selection step involves either binding of the desired protein to an immobilized binding partner or assaying for a functional activity of the desired protein. In addition, the method may further involve repeating steps (a) and (b).

In a final aspect, the invention features a solid support including an array of immobilized molecules, each including a covalently-bonded DNA-protein fusion of the invention. In a preferred embodiment, the solid support is a microchip.

As used herein, by a "population" is meant $10^5$ or more molecules

(for example, DNA-protein fusion molecules). Because the methods of the invention facilitate selections which begin, if desired, with large numbers of candidate molecules, a "population" according to the invention preferably means more than $10^7$ molecules, more preferably, more than $10^9$, $10^{13}$, or $10^{14}$ molecules, and, most preferably, more than $10^{15}$ molecules.

By "selecting" is meant substantially partitioning a molecule from other molecules in a population. As used herein, a "selecting" step provides at least a 2-fold, preferably, a 30-fold, more preferably, a 100-fold, and, most preferably, a 1000-fold enrichment of a desired molecule relative to undesired molecules in a population following the selection step. A selection step may be repeated any number of times, and different types of selection steps may be combined in a given approach.

By a "protein" is meant any two or more naturally occurring or modified amino acids joined by one or more peptide bonds. "Protein" and "peptide" are used interchangeably herein.

By "RNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified ribonucleotides. One example of a modified RNA included within this term is phosphorothioate RNA.

By "DNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified deoxyribonucleotides.

By a "nucleic acid" is meant any two or more covalently bonded nucleotides or nucleotide analogs or derivatives. As used herein, this term includes, without limitation, DNA, RNA, and PNA.

By a "peptide acceptor" is meant any molecule capable of being added to the C-terminus of a growing protein chain by the catalytic activity of the ribosomal peptidyl transferase function. Typically, such molecules contain (i) a nucleotide or nucleotide-like moiety (for example, adenosine or an adenosine analog (di-methylation at the N-6 amino position is acceptable)), (ii)



an amino acid or amino acid-like moiety (for example, any of the 20 D- or L-amino acids or any amino acid analog thereof (for example, O-methyl tyrosine or any of the analogs described by Ellman et al., Meth. Enzymol. 202:301, 1991), and (iii) a linkage between the two (for example, an ester, amide, or ketone linkage at the 3' position or, less preferably, the 2' position); preferably, this linkage does not significantly perturb the pucker of the ring from the natural ribonucleotide conformation. Peptide acceptors may also possess a nucleophile, which may be, without limitation, an amino group, a hydroxyl group, or a sulfhydryl group. In addition, peptide acceptors may be composed of nucleotide mimetics, amino acid mimetics, or mimetics of the combined nucleotide-amino acid structure.

By an "altered function" is meant any qualitative or quantitative change in the function of a molecule.

By "binding partner," as used herein, is meant any molecule which has a specific, covalent or non-covalent affinity for a portion of a desired DNA-protein fusion. Examples of binding partners include, without limitation, members of antigen/antibody pairs, protein/inhibitor pairs, receptor/ligand pairs (for example cell surface receptor/ligand pairs, such as hormone receptor/peptide hormone pairs), enzyme/substrate pairs (for example, kinase/substrate pairs), lectin/carbohydrate pairs, oligomeric or heterooligomeric protein aggregates, DNA binding protein/DNA binding site pairs, RNA/protein pairs, and nucleic acid duplexes, heteroduplexes, or ligated strands, as well as any molecule which is capable of forming one or more covalent or non-covalent bonds (for example, disulfide bonds) with any portion of a DNA-protein fusion.

By a "solid support" is meant, without limitation, any column (or column material), bead, test tube, microtiter dish, solid particle (for example, agarose or sepharose), microchip (for example, silicon, silicon-glass, or gold

chip), or membrane (for example, the membrane of a liposome or vesicle) to which an affinity complex may be bound, either directly or indirectly (for example, through other binding partner intermediates such as other antibodies or Protein A), or in which an affinity complex may be embedded (for example, through a receptor or channel).

The present invention provides methods for the creation of fusions between proteins and their encoding cDNAs. These constructs possess greatly enhanced chemical stability, first, due to the DNA component of the fusion and, second, due to the covalent bond linking of the DNA and protein moieties. These properties allow for easier handling of the fusion products and thereby allow selection and recognition experiments to be carried out under a range of reaction conditions. In addition, the present invention facilitates applications where a single-stranded nucleic acid portion is mandatory, for example, in hybridization assays in which the coding fusions are immobilized to a solid support. In addition, incubations may be performed under more rigorous conditions, involving high pH, elevated concentrations of multivalent metal ions, prolonged heat treatment, and exposure to various biological materials. Finally, single-stranded DNA is relatively resistant to secondary structure formation, providing a great advantage for techniques involving or requiring nucleic acid hybridization steps.

In addition, the methods of the present invention allow for the production of fusions involving DNA and protein components of any length, as well as fusion libraries of high complexity.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

## Brief Description of the Drawings

FIGURE 1 is a schematic illustration of a method for the generation



of DNA-protein fusions (Type A1) that involves ligation of a puromycin-modified DNA hairpin-like structure to an mRNA molecule.

FIGURE 2 is a schematic illustration of a method for the generation of branched hairpin structures.

FIGURE 3 is a schematic illustration of a method for the synthesis of puromycin-5'-phosphoramidite.

FIGURE 4 is a schematic illustration of a method for the generation of branched hairpin structures.

FIGURE 5 is a schematic illustration of a method for the generation of DNA-protein fusions that involves photocrosslinking of a 5'-psoralen-modified primer DNA to a suitable linker that bears a 3'-puromycin.

FIGURE 6 is a schematic illustration of exemplary methods for the chemical ligation of mRNA and DNA molecules.

FIGURE 7 is a schematic illustration of a method for the synthesis of hydrazide phosphoramidite.

FIGURE 8 is a schematic illustration of a method for the synthesis of hydrazine phosphoramidite.

FIGURE 9 is a schematic illustration of a method for the generation of DNA-protein fusions that involves chemical crosslinking of a puromycin-modified linker to the 3'-end of an mRNA molecule.

FIGURE 10 is a schematic illustration of a method for the generation of DNA-protein fusions that involves psoralen-mediated photocrosslinking of a combined linker/reverse transcription primer construct to the 3'-end of an mRNA molecule.

FIGURE 11 is a schematic illustration of an alternative method for the generation of DNA-protein fusions that involves psoralen photocrosslinking of a combined linker/reverse transcription primer construct.

FIGURE 12 is a schematic illustration of a method for the generation

of DNA-protein fusions that involves crosslinking of a reverse transcription primer to a preexisting mRNA-linker construct.

FIGURE 13 is a schematic illustration of a method for the generation of DNA-protein fusions that involves crosslinking of a reverse transcription primer to a preexisting mRNA-protein fusion.

FIGURE 14 is a schematic illustration of the oligonucleotide constructs (SEQ ID NOS: 1-6) used for the preparation of the exemplary DNA-protein fusions described herein.

FIGURE 15 is a schematic illustration of the preparation of Type C2 DNA-protein fusions.

FIGURE 16 is a photograph illustrating a product analysis of the Type C2 DNA-protein fusions.

FIGURE 17 is a schematic illustration of the preparation of Type B3 DNA-protein fusions.

FIGURE 18 is a schematic illustration of the preparation of Type B2 DNA-protein fusions.

FIGURE 19 is a photograph illustrating the resistance analysis of Type B3 DNA-protein fusions against nuclease and base treatment.

FIGURE 20 is a graph illustrating the experimentally determined half-lives of RNA- and DNA-protein fusion products in the presence of cell membrane preparations.

## Detailed Description

There are now provided below a number of exemplary techniques for the production of DNA-protein fusions, and descriptions for their use. These examples are provided for the purpose of illustrating, and not limiting, the invention.



### Type A1: Template-Directed Ligation of a Puromycin-Modified Hairpin-Like Structure to an mRNA

According to a first exemplary approach, DNA-protein fusions are generated by ligating a puromycin-modified DNA hairpin-like structure to an mRNA molecule, as illustrated in Figure 1. The first step of this procedure is the attachment of puromycin to the hairpin, and this may be accomplished by a number of techniques, one of which is shown in Figure 2. By this approach, a DNA hairpin is synthesized with a puromycin-terminated side chain branching out from the DNA molecule. This construct may be generated using an asymmetric branched phosphoramidite (Clontech, Palo Alto, CA) in any standard automated DNA synthesis of the hairpin structure (see, for example, User Guide to Expedite Nucleic Acid Synthesis System, Perseptive Biosystems, Framingham, MA), followed by the addition of a 5'-phosphate using a chemical phosphorylation reagent (Glen Research, Sterling VA).

Subsequently the protecting group is selectively removed from the branch (Product Protocol for Asymmetric Branching Phosphoramidite, Clontech, Palo Alto, CA), followed by the attachment of the linker portion through standard automated DNA synthesis. Before reaching the end of the linker, the strand orientation is reversed by the addition of a few 5'-phosphoramidites (Glen Research, Sterling, VA). Finally, the synthesis is terminated through attachment of the puromycin-5'-phosphoramidite, preferably using the synthetic technique shown in Figure 3. In Figure 3, steps (a)-(c) may be carried out as described in Greene & Wuts (Protective Groups in Organic Synthesis, 2nd ed. (1991) John Wiley & Sons, Inc., New York, New York), and step (d) may be carried out as described in Beaucage (Methods in Molecular Biology, vol. 20, Protocols for Oligonucleotides and Analogs, ed. S. Agarwal (1993) Humana Press, Totowa, N.J., pp. 33-61).

Alternatively, the puromycin-modified branched hairpin may be

synthesized as shown in Figure 4. By this technique, synthesis is initiated from a puromycin-CPG solid support (Glen Research, Sterling, VA) by first synthesizing the linker portion, followed by incorporation of the branched amidite (Clontech, Palo Alto, CA) and addition of the 5'-portion of the hairpin. After deprotection of the branch, the 3'-arm of the hairpin is added by using nucleoside-5'-
phosphoramidites (Glen Research, Sterling, VA).

By either of the above approaches, in the next step, the mRNA is ligated to the hairpin, for example, using T4 DNA ligase and the 3'-overhang as a template (Sambrook, Fritsch & Maniatis Molecular Cloning (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). Ribosomal translation of the RNA then leads to protein synthesis with subsequent fusion formation (see, for example, Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302). In one particular embodiment, the branching point is located in the loop region of the hairpin. Other positions of the branching point (e.g., within the stem structure) may also be utilized. In addition, while a $dA_n$ linker of between approximately 10-60 nucleotides, and more preferably approximately 30 nucleotides, is utilized, both the length and the chemical composition (e.g., PEG (Glen Research, Sterling, VA) rather than $dA_n$) of the linker may be optimized.

In a final step, the RNA portion of the construct is reverse transcribed into cDNA (for example, as described in Sambrook, Fritsch & Maniatis, Molecular Cloning, (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) using the hairpin 3' end as a primer. Optional digestion of the mRNA by RNase H (see, for example, Sambrook, Fritsch & Maniatis Molecular Cloning, (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) yields a single stranded DNA-protein fusion.



This method also facilitates the formation of truncated DNA transcripts by adding didesoxynucleoside triphosphates during transcription (see, for example, Sanger, Science (1981) vol. 214, p. 1205-1210). Such truncated DNA-protein fusions are useful in protein display experiments (Kumelis et al., U.S.S.N. 60/080,686, filed April 3, 1998), for example, where only the 3'-region of the original message (now the 5'-region of the DNA transcript) is used for hybridization with immobilized oligonucleotide probes.

Type A2: Crosslinking of a Puromycin-Modified Linker to a Primer DNA

As an alternative to the hairpin-like construct described above, a closely related structure may also be prepared through photocrosslinking of a 5'-psoralen-modified primer DNA with a suitable linker that bears a 3'-puromycin. An exemplary crosslinking method is illustrated in Figure 5. In this method, the puromycin-bearing linker may be constructed as described, for example, in Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. The psoralen-modified primer may be generated and the photocrosslinking step carried out as described, for example, in Pieles & Englisch, Nucl. Acids Res. (1989) vol. 17, p. 285-299. The remaining steps may be carried out as described above. This approach does not require the use of non-standard nucleoside/ puromycin-5'-phosphoramidites (i.e., which were used during the automated synthesis of the hairpin-linker structure), providing an advantage over the hairpin method. Again, as above, while a $dA_n$ linker of between approximately 10-60 nucleotides, and more preferably approximately 30 nucleotides, is utilized, both the length and the chemical composition (e.g., PEG (Glen Research, Sterling, VA) rather than $dA_n$) of the linker may be optimized.

In addition, for each of the Type A1 and Type A2 methods, the

ligation reaction between the mRNA and the DNA portion of the construct may be carried out by several alternative techniques. For example, in addition to the enzymatic ligation with T4 DNA ligase described above, this step may be accomplished using chemical methods. In one particular example, the 5'-end of the hairpin may be modified with one (or multiple) amino-groups using the appropriate phosphoramidite (Clontech, Palo Alto, CA). After periodate oxidation of the 3'-end of the RNA, the two substrates may be joined through a reductive amination reaction. This is illustrated as scheme "A" in Figure 6 and is described, for example, in Lemaitre et al., Proc. Natl. Acad. Sci. USA (1987) vol. 84, p. 648-652. Alternatively, this chemical ligation step may involve carbohydrazide or hydrazine modified structures for hydrazone formation or reductive amination. These approaches are illustrated in Figure 6, respectively, as schemes "B" and "C" and are described, respectively, in Gosh et al. (Anal. Biochem. (1989) vol. 178, p. 43-51) and Proudnikov & Mirzabekov (Nucl. Acids Res. (1996) vol. 24 p. 4535-4542). Hydrazide phosphoramidite synthesis may be carried out as shown in Figure 7, and hydrazine phosphoramidite synthesis as shown in Figure 8 and as described in Greene & Wuts (Protective Groups in Organic Synthesis, 2nd ed. (1991) John Wiley & Sons, Inc., New York, New York (steps (a) and (c)), Proudnikov & Mirzabekov (Nucl. Acids Res. (1996) vol. 24 p. 4535-4542 (step b)), and Beaucage (Methods in Molecular Biology, vol. 20, Protocols for Oligonucleotides and Analogs, ed. S. Agarwal (1993) Humana Press, Totowa, N.J., pp. 33-61 (step (e)).

Types B1-B3: Chemical Crosslinking to the 3'-end of an mRNA

Yet another approach to the generation of DNA-protein fusions involves the chemical crosslinking of a puromycin-modified linker to the 3'-end of an mRNA molecule. Such crosslinking may be accomplished by a number of approaches.



One exemplary approach is shown schematically in Figure 9. In this approach ("B1"), an oligonucleotide is synthesized that bears a reactive group (e.g., one of the amino derivatives, hydrazides, or hydrazines described above) located between the primer and the linker regions. Duplex formation of the RNA and the primer site takes place immediately adjacent to this reactive group, which then is allowed to react with the periodate-oxidized 3'-end of the RNA leading to a crosslink (as shown in Figure 6 and as described above). This reaction may occur through reductive amination (Figure 6, scheme "A" or "C"; Lemaitre et al., Proc. Natl. Acad. Sci. USA (1987) vol. 84, p. 648-652; Proudnikov & Mirzabekov, Nucl. Acids Res. (1996) vol. 24 p. 4535-4542) or hydrazone formation (Figure 6, scheme "B"; Gosh et al., Anal. Biochem. (1989) vol. 178, p. 43-51). Following translation and fusion formation (Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302), the primer is extended by reverse transcriptase on the RNA template and an optional RNase H digestion step is carried out, generating the DNA-protein fusion (Figure 9).

As in methods A1 and A2 above, the strand direction of the linker portion's terminal nucleotides is reversed, which can be accomplished by the use of 5'-phosphoramidites (Glen Research, Sterling, VA) during synthesis.

In yet another exemplary crosslinking approach ("B2"), a photoreactive psoralen moiety is included in the linker as a reactive group (Figure 10). Such a construct may be synthesized using a psoralen-modified desoxynucleotide phosphoramidite (Pieles et al., Nucleic Acids Res. (1989) vol. 17, p. 8967-8978) or by incorporating a branched phosphoramidite (Clontech, Palo Alto, CA) to which a standard psoralen phosphoramidite is attached (Glen Research, Sterling, VA). Following hybridization of the linker to the target RNA, crosslink formation is achieved through irradiation with UV-light, for

example, as described in Pieles and Englisch (Nucl. Acids Res. (1989) vol. 17, p. 285-299). The resulting construct is then subjected to translation and fusion formation (Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO 98/31700; Roberts and Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302). Reverse transcription and RNase H digestion yields the final DNA-protein fusions.

Alternatively, crosslinking may be accomplished using a combined linker/reverse transcriptase primer construct as depicted in Figure 11 ("B3"). In a variant of the above approach, the psoralen moiety is not directly attached between the linker and primer region, but rather connected to a short DNA branch. This DNA portion also hybridizes to the target RNA and thus provides an optimized double-stranded environment for the psoralen to react (Pieles and Englisch, Nucl. Acids. Res. (1989) vol. 17, p. 285-299). Preparation of DNA-protein fusions using this psoralen construct may be carried out as described above.

Types C1 and C2: Crosslinking of the Reverse Transcription Primer to Preexisting mRNA-Linker Constructs

Another method for generating DNA-protein fusions is shown schematically in Figure 12. By this approach, RNA is initially ligated to a linker molecule as previously described (Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302). In a subsequent step, a suitable primer bearing a 5'-photocrosslinking reagent (e.g., psoralen, Glen Research, Sterling, VA) is annealed to the RNA-linker product. Irradiation with light furnishes a covalent crosslink between the two oligonucleotide strands (as described, for example, in Pieles & Englisch, Nucl. Acids Res. (1989) vol. 17, p. 285-299). As in methods Type A1, A2, and B1-



B3 above, translation and fusion formation may be carried out, followed by a reverse transcription step and an optional RNase H digestion step to yield DNA-protein fusions (Figure 12).

Alternatively, as shown in Figure 13, the initial steps of the above procedure may be carried out in the opposite order. This approach allows translation and fusion formation to be performed prior to crosslinking and reverse transcription. Accordingly, this method allows for the use of previously described and well established reaction conditions and components for translation and RNA-protein fusion formation.

Experimental Results

Exemplary techniques described above were carried out to demonstrate DNA-protein fusion formation. These experiments made use of the oligonucleotides depicted in Figure 14.

Model RNA substrates 1: GGG ACA AUU ACU AUU UAC AAU UAC AAU GGA CUA CAA GGA CGA UGA CGA UAA GGG CGG CUG GUC CCA CCC CCA GUU CGA GAA GGC AUC CGC U (SEQ ID NO: 7); 2: GGG ACA AUU ACU AUU UAC AAU UAC AAU GGA CUA CAA GGA CGA UGA CGA UAA GGG CGG CUG GUC CCA CCC CCA GUU CGA GAA GGC AUC CGC UCU UUC ACU AUA (SEQ ID NO: 8); and 3: GGG ACA AUU ACU AUU UAC AAU UAC AAU GGA CUA CAA GGA CGA UGA CGA UAA GGG CGG CUG GUC CCA CCC CCA GUU CGA GAA GGC AUC CGC UAU UUA AAA AAA AAA AAA AAA AAA A (SEQ ID NO: 9) were synthesized by T7 transcription (Megashortscript transkiption kit, Ambion, Austin, TX) using appropriate dsDNA templates. Following transcription, the RNAs were purified by denaturing polyacrylamide gel electrophoresis.

The modified oligonucleotides 4: 5' pd(AAA AAA AAA ACG GCT

ATA TAA AAA AAA CC)- Pu (SEQ ID NO: 10); **5**: 5' psoralen C2-TAG CCG TTT TTT TTT TAG CGG ATG C (SEQ ID NO: 11); **6**: 5' d(cgt agg cga gaa agt gat)-branch[psoralen C6]-d(AAA AAA AAA AAA AAA AAA AAA AAA AAA CC)-Pu (SEQ ID NO: 12); and **7**: 5' ggt caa gct ctt-branch[5' psoralen C6-TAG CGG ATG C 3'] $spacer_6$ CC-Pu (SEQ ID NO: 13) [[uppercase = standard DNA-3'-phosphoramidites; lowercase = DNA-5'-phosphoramidites; spacer = spacer-9 phosphoramidite; Pu = puromycin-CPG (all from Glen Research, Sterling, VA); branch = asymmetric branching amidite (Clontech, Palo Alto, CA)] were synthesized on an Expedite Synthesizer Model 8909 (PerSeptive Biosystems, Framingham, MA) according to recommended protocols for the corresponding phosphoramidites. For the branched constructs **6** and **7**, the main chain was synthesized first and concluded with a final capping step. Next, the levulinyl protecting group was removed from the branching unit through treatment with 0.5 M hydrazine monohydrate in pyridine-acetic acid for 15 minutes at room temperature. Automated synthesis was then resumed and the side chain sequences (indicated in square brackets) were attached. The oligos were fully deprotected in concentrated ammonium hydroxide for 8 hours at 55°C and purified by denaturing gel electrophoresis.

The DNA sequences **8**: d(TTT TTT TTT TAG CGG ATG C) (SEQ ID NO: 14) and **9**: d(TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT) (SEQ ID NO: 15) were purchased from Oligos etc. (Wilsonville, OR) and used without further purification.

Type C2 DNA-Protein Fusion Formation

Type C2 DNA-protein fusion formation was demonstrated as follows (Figure 15). RNA **1** and linker **4** were hybridized to template DNA **8** and enzymatically ligated by T4 DNA ligase as previously described (Szostak et al.,



U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts and Szostak, Proc. Natl. Acad. Sci. USA (1997) Vol. 94, p. 12297-12302). After purification by electrophoresis on a denaturing polyacrylamide gel, the resulting mRNA-linker construct was used as a template for *in vitro* translation using rabbit reticulocyte lysate kits from Ambion. Reaction mixtures contained 50 pmole ligated mRNA **10**, 10 mM creatine phosphate, 150 mM potassium acetate, 0.5 mM magnesium chloride, 0.1 mM of each amino acid exept methionine, 150 µCi [$^{35}$S] methionine (Amersham, Arlington Heights, IL) and 67% v/v of lysate in a total volume of 300 µl and were carried out for 30 minutes at 30°C. To promote the subsequent fusion formation, KCl and $MgCl_2$ were added to 590 mM and 50 mM final concentrations, respectively, in a volume of 500 µl. Incubation was continued for 60 minutes at 20°C. Products were isolated by diluting the lysate into 10 ml of binding buffer (100 mM Tris pH 8.0, 10 mM EDTA, 1 M NaCl, 0.25% v/v Triton X-100) and adding 10 mg oligo-dT cellulose type 7 (Pharmacia, Piscataway, NJ). Samples were rotated for 60 minutes at 4°C, and the solid support was then washed with 5 ml ice-cold binding buffer that was devoid of EDTA, followed by elution with 100 µl aliquots of water. Fusion product was found in fractions 2 and 3, and these fractions were combined. The total yield of fusion **11** was determined by scintillation counting of the incorporated [$^{35}$S] methionine to be 1.6 pmole (3.2% of input RNA).

For the conversion of the RNA-protein fusions **11** into DNA-protein fusions **13**, the following reactions were performed (Figure 15). First, 20 µl of the above oligo-dT-purified material **11** was mixed with 0.5 µl primers **5** (50 µM) and 6 µl first strand buffer (Superscript II kit from GibcoBRL; 250 mM Tris-HCl pH 8.3, 375 KCl, 15mM $MgCl_2$) and briefly heated to 80°C for 2 minutes, followed by slowly cooling to 0°C. Psoralen photocrosslink formation was induced by irradiating the sample for 15 minutes at 0°C with $\lambda >$

310 nm [450 W medium pressure immersion lamp (ACE Glass, Vineland, NJ) equipped with a Pyrex absorption sleeve in a Quartz immersion well]. Next, 0.6 µl of a dNTP mix (25 mM each), 3 µl of 0.1M DTT, and 0.4 µl (80 units) Superscript II reverse transcriptase were added, and cDNA synthesis was carried out for 60 minutes at 42°C. The RNA portion was then removed by continuing incubation for 60 minutes at 37°C after addition of 0.5 µl (1 unit) RNase H (Promega, Madison, WI). Finally, double-stranded DNA **14** was generated by adding 50 pmoles of primer **9** and incubating for another 60 minutes at 42°C. Control reactions with non-crosslinked samples were performed as indicated in Figure 15. Product analysis was performed by electrophoresis on denaturing 6% TBE-Urea gels (Novex, San Diego, CA), followed by visualization of the [$^{35}$S]-labelled product bands by exposure on a phosphorimager screen (Figure 16).

Samples were applied to the gel in the same order as they appear in Figure 15, beginning with RNA-protein fusion **11** and following the reaction pathway with and without having been photocrosslinked. As indicated in Figure 16, the gel mobilities correspond well with the expected behavior and clearly confirm the constitution of DNA-protein fusion **13**.

Type B3 DNA-Protein Fusion Formation

Type B3 DNA-protein fusion formation was demonstrated as follows (Figure 17). The branched linker construct **7** (5 µM) was annealed to the target RNA **3** (2.5 µM) in 25 mM Tris buffer pH 7.0 containing 100 mM NaCl and crosslinked by irradiation for 15 minutes at room temperature in a borosilicate glass vial (Kimble/Kontes, Vineland, NJ) using a handheld multiwavelength UV lamp model UVGL-25 (UVP, Upland, CA) set to long wave. Product analysis was performed by electrophoresis on a 6% TBE-Urea polyacrylamide gel followed by visualization by UV shadowing. These results indicated nearly



quantitative conversion of the starting material (gel "A" in Fig. 17). The photoligated product RNA was used for *in vitro* translation without further separation from remaining unligated RNA and excess linker. *In vitro* translation and fusion formation reactions were performed as described for Type C2 above, with 100 pmole input RNA in a 300 µl total volume. After purification on oligo-dT cellulose, 5.5 pmole RNA-fusion **15** was obtained. Its conversion into single-stranded and double-stranded DNA-protein fusions **16** and **17**, respectively, was done by reverse transcription (Superscript II kit, GibcoBRL, Grand Island, NY) and RNase H (Promega, Madison, WI) treatment as described for Type C2 fusions (gel "B" in Fig. 17).

Type B2 DNA-Protein Fusion Formation

Type B2 DNA-protein fusion formation was demonstrated as outlined in Figure 18. Specifically, following the procedure outlined for Type B3 fusions above, RNA **2** was crosslinked to linker **6**. Following denaturing polyacrylamide electrophoresis, the ligated product **18** was isolated in 12% yield. *In vitro* translation, fusion formation, and preparation of DNA-protein fusions **19** were carried out as described for Type B3 fusions above, with similar efficiencies of fusion formation.

DNA-Protein Fusion Stability Tests

To evaluate the nuclease and base resistance of DNA fusions in comparison with the corresponding RNA fusions, the following experiments were carried out. To 10 µl DNA-fusion **16** (Type B3) or RNA-fusion **15** in reverse transcription buffer was added either 0.2 µl (0.4 units) RNase H, 0.2 µl (2 units) RNase I, 0.2 µl (0.6 units) T4 DNA polymerase (3'-5' exonuclease activity), or 2.5 µl of 2.0 M NaOH. Samples were incubated for 30 minutes at 37°C and then analyzed on a 4-12% NuPage gel (Novex, San Diego, CA)

followed by autoradiography. Results are shown in Figure 19 and confirm the increased stability of DNA fusions against ribonucleases and base treatment.

To test stability of DNA fusion constructs in biological media, 5 nM of either RNA fusions **11** or **12**, or DNA fusions **13** or **14** (Type C2) were incubated with 3 µg/µl CHO-K1 cell membranes (Receptor Biology, Beltsville, MD) in 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, and 10 mM DTT at room temperature. Additional samples of RNA fusions **11** and **12** were prepared containing 20 mM vanadyl ribonucleside complex ("VRC") to inhibit ribonuclease activity. Aliquots were taken after 0, 5, 15, 30, 60, 120 minutes, and 24 hours and analyzed by electrophoresis on 4-12% NuPage gels (Novex) followed by exposure on a phosphorimager screen. The relative amounts of remaining fusion were plotted against incubation time and half-lives graphically extracted from the resulting curves. As indicated in Figure 20, all constructs showed more than 50% decay during the initial two hour period except for dsDNA fusion **14**, which appeared to be entirely stable under the conditions tested. Following a 24 hour incubation, all fusion constructs were completely degraded due to either nuclease or protease activity.

In Vitro Selection of Desired Proteins

The DNA-protein fusions described herein may be used in any selection method for desired proteins, including molecular evolution and recognition approaches. Exemplary selection methods are described, for example, in Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302; Lipovsek et al., U.S.S.N. 60/096,818 and U.S.S.N. 09/374,962; and Kuimelis et al. U.S.S.N. 60/080,686 and U.S.S.N. 09/282,734, all hereby incorporated by reference.



Use

The DNA-protein fusions described herein may be used for any application previously described or envisioned for RNA-protein fusions. Commercial uses include the isolation of polypeptides with desired properties through *in vitro* evolution techniques (see, for example, Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302)), screening of cDNA libraries that are derived from cellular mRNA (see, for example, Lipovsek et al., U.S.S.N. 60/096,818, filed August 17, 1998), and the cloning of new genes on the basis of protein-protein interactions (Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09,247,190; Szostak et al., WO98/31700), as well as the use of these fusions in protein display experiments (Kuimelis et al. U.S.S.N. 60/080,686 and U.S.S.N. 09/282,734). In addition, the DNA-protein fusions described herein may be used in binding and molecular recognition assays that involve biological materials that presumably contain ribonucleases, such as whole cells, lysates, or biological fluids. These DNA-protein fusions may be used for any appropriate therapeutic, diagnostic, or research purpose, particularly in the pharmaceutical and agricultural areas.

What is claimed is:

Claims

1. A method for generating a DNA-protein fusion, said method comprising:

(a) linking a nucleic acid primer to an RNA molecule, said primer being bound to a peptide acceptor;

(b) translating said RNA to produce a protein product, said protein product being covalently bound to said primer; and

(c) reverse transcribing said RNA to produce a DNA-protein fusion.

2. A method for generating a DNA-protein fusion, said method comprising:

(a) generating an RNA-protein fusion;

(b) hybridizing a nucleic acid primer to said fusion; and

(c) reverse transcribing said RNA to produce a DNA-protein fusion.

3. The method of claim 1 or 2, said method further comprising treating the product of step (c) to remove said RNA.

4. The method of claim 3, wherein said treating comprises contacting the product of step (c) with RNase H under conditions sufficient to digest said RNA.

5. The method of claim 1 or 2, wherein said nucleic acid primer is a DNA primer.

6. The method of claim 1 or 2, wherein said translating step is carried out in vitro.





7. The method of claim 1, wherein said peptide acceptor is puromycin.

8. The method of claim 1, wherein said nucleic acid primer has a hairpin structure.

9. The method of claim 1 or 2, wherein said primer further comprises a photocrosslinking agent.

10. The method of claim 9, wherein said photocrosslinking agent is psoralen.

11. The method of claim 9, wherein said primer is crosslinked to an oligonucleotide which is bound to a peptide acceptor.

12. The method of claim 9, wherein said primer is hybridized to said RNA molecule and said linking step is carried out by photocrosslinking.

13. A molecule comprising a DNA covalently bonded to a protein through a peptide acceptor.

14. The molecule of claim 13, wherein said peptide acceptor is puromycin.

15. The molecule of claim 13, wherein said protein comprises at least 10 amino acids.

16. A molecule comprising a DNA covalently bonded to a protein,

said protein comprising at least 10 amino acids.

17. The molecule of claim 13 or 16, wherein said protein comprises at least 30 amino acids.

18. The molecule of claim 17, wherein said protein comprises at least 100 amino acids.

19. The molecule of claim 13 or 16, wherein said protein is encoded by said DNA.

20. The molecule of claim 13 or 16, wherein said protein is entirely encoded by said DNA.

21. The molecule of claim 13 or 16, wherein said molecule further comprises a ribonucleic acid covalently bonded to said DNA.

22. The molecule of claim 21, wherein said protein is encoded by said ribonucleic acid.

23. The molecule of claim 13 or 16, wherein said DNA is double stranded.

24. A population of at least $10^5$ DNA-protein fusions, each fusion comprising a DNA covalently bonded to a protein.

25. The population of claim 24, wherein said population comprises at least $10^{14}$ fusions.



26. The population of claim 24, wherein said protein comprises at least 10 amino acids.

27. The population of claim 26, wherein said protein comprises at least 30 amino acids.

28. The population of claim 27, wherein said protein comprises at least 100 amino acids.

29. The population of claim 24, wherein said protein is encoded by said covalently bonded DNA.

30. The population of claim 24, wherein said protein is entirely encoded by said covalently bonded DNA.

31. The population of claim 24, wherein said fusions further comprise a ribonucleic acid covalently bonded to said DNA.

32. The population of claim 31, wherein said protein is encoded by said ribonucleic acid.

33. The population of claim 24, wherein said DNA is covalently bonded to said protein through a peptide acceptor.

34. The population of claim 33, wherein said peptide acceptor is puromycin.

35. The population of claim 24, wherein said DNA is double

stranded.

36. A method for the selection of a desired protein or its encoding DNA, comprising the steps of:

a) providing a population of DNA-protein fusions, each comprising a DNA covalently bonded to a candidate protein; and

b) selecting a desired DNA-protein fusion, thereby selecting said desired protein or DNA.

37. A method for the selection of a protein, or its encoding DNA, having an altered function relative to a reference protein, comprising the steps of:

a) producing a population of candidate DNA-protein fusions, each comprising a DNA covalently bonded to a candidate protein and having a candidate protein coding sequence which differs from said reference protein coding sequence; and

b) selecting a DNA-protein fusion having an altered function, thereby selecting said protein having said altered function or its encoding DNA.

38. The method of claim 36 or 37, wherein said protein is encoded by said DNA.

39. The method of claim 36 or 37, wherein said DNA-protein fusion further comprises a ribonucleic acid covalently bonded to said DNA.

40. The method of claim 39, wherein said protein is encoded by said ribonucleic acid.



41. The method of claim 36 or 37, wherein said protein comprises at least 10 amino acids.

42. The method of claim 41, wherein said protein comprises at least 30 amino acids.

43. The method of claim 42, wherein said protein comprises at least 100 amino acids.

44. The method of claim 36 or 37, wherein said DNA is covalently bonded to said protein through a peptide acceptor.

45. The method of claim 44, wherein said peptide acceptor is puromycin.

46. The method of claim 36 or 37, wherein said population of candidate DNA-protein fusions comprises at least $10^5$ different DNA molecules.

47. The method of claim 46, wherein said population of candidate DNA-protein fusions comprises at least $10^{14}$ different DNA molecules.

48. The method of claim 36 or 37, wherein said selection step comprises binding of said desired protein to an immobilized binding partner.

49. The method of claim 36 or 37, wherein said selection step comprises assaying for a functional activity of said desired protein.

50. The method of claim 36 or 37, wherein said method further comprises repeating steps (a) and (b).

51. The method of claim 36 or 37, wherein said DNA is double stranded.

52. A solid support comprising an array of immobilized molecules, each comprising a DNA covalently bonded to a protein.

53. The solid support of claim 52, wherein said protein is encoded by said DNA.

54. The solid support of claim 52, wherein said solid support is a microchip.

WO 00/32823 PCT/US99/28472



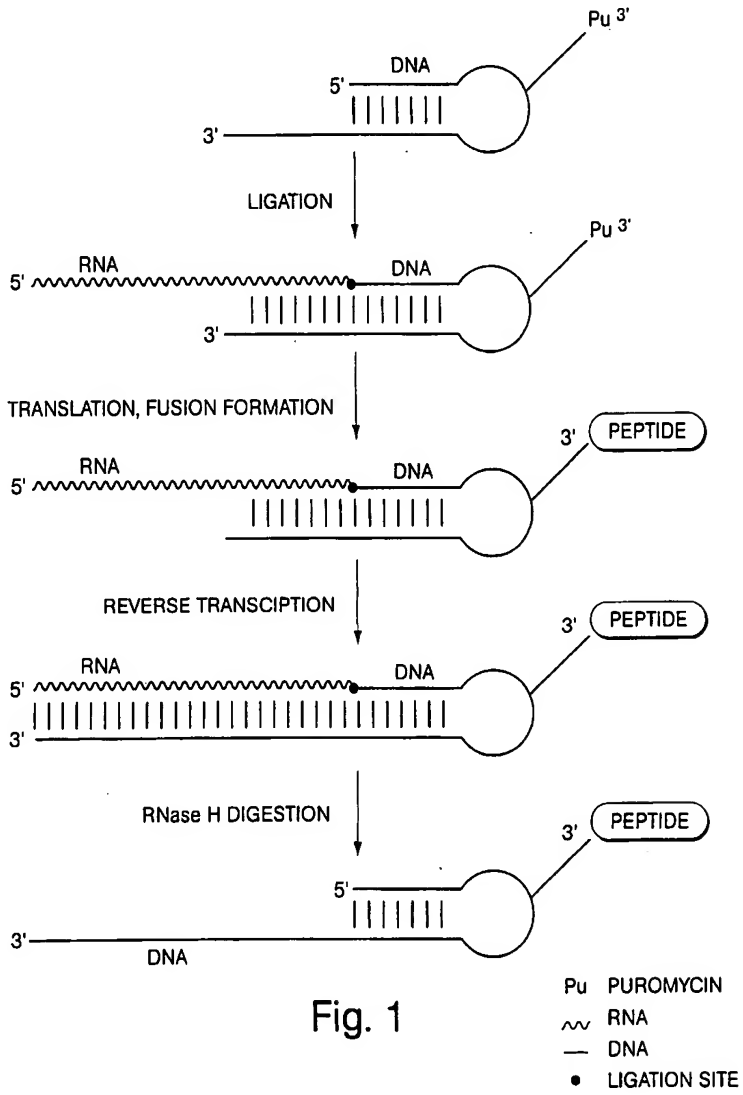



Fig. 1

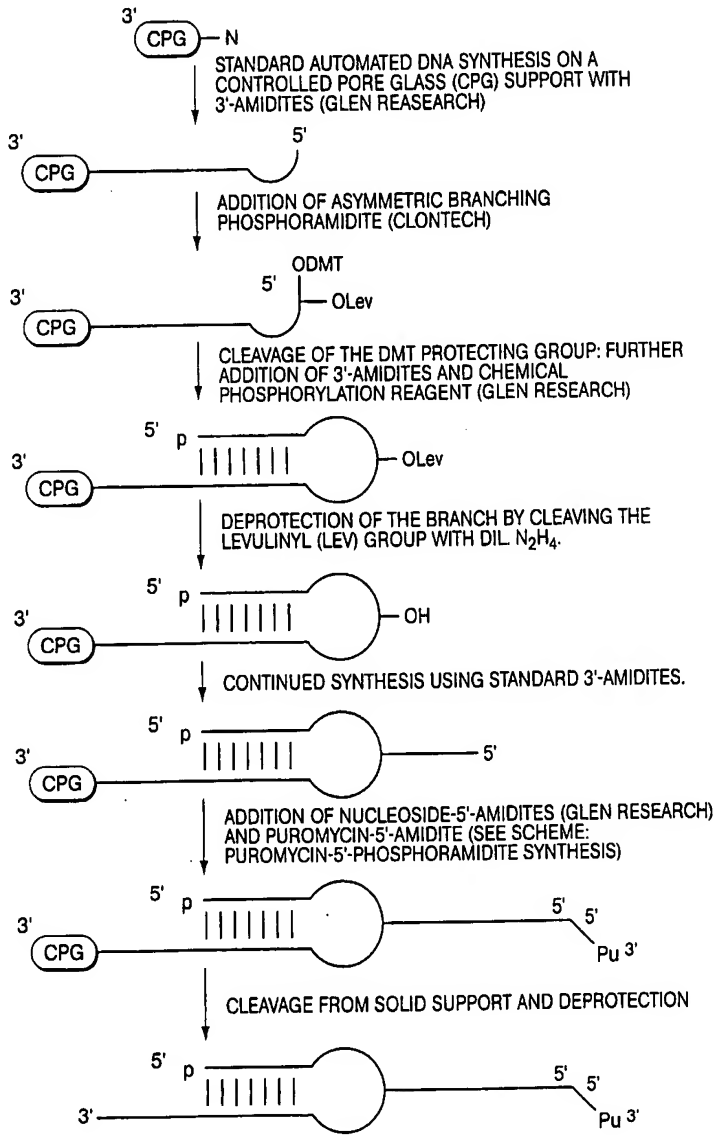

3'
CPG
N
STANDARD AUTOMATED DNA SYNTHESIS ON A CONTROLLED PORE GLASS (CPG) SUPPORT WITH 3'-AMIDITES (GLEN REASEARCH)
5'
ADDITION OF ASYMMETRIC BRANCHING PHOSPHORAMIDITE (CLONTECH)
ODMT
OLev
CLEAVAGE OF THE DMT PROTECTING GROUP: FURTHER ADDITION OF 3'-AMIDITES AND CHEMICAL PHOSPHORYLATION REAGENT (GLEN RESEARCH)
p
DEPROTECTION OF THE BRANCH BY CLEAVING THE LEVULINYL (LEV) GROUP WITH DIL. $N_2H_4$.
OH
CONTINUED SYNTHESIS USING STANDARD 3'-AMIDITES.
ADDITION OF NUCLEOSIDE-5'-AMIDITES (GLEN RESEARCH) AND PUROMYCIN-5'-AMIDITE (SEE SCHEME: PUROMYCIN-5'-PHOSPHORAMIDITE SYNTHESIS)
Pu 3'
CLEAVAGE FROM SOLID SUPPORT AND DEPROTECTION

Fig. 2







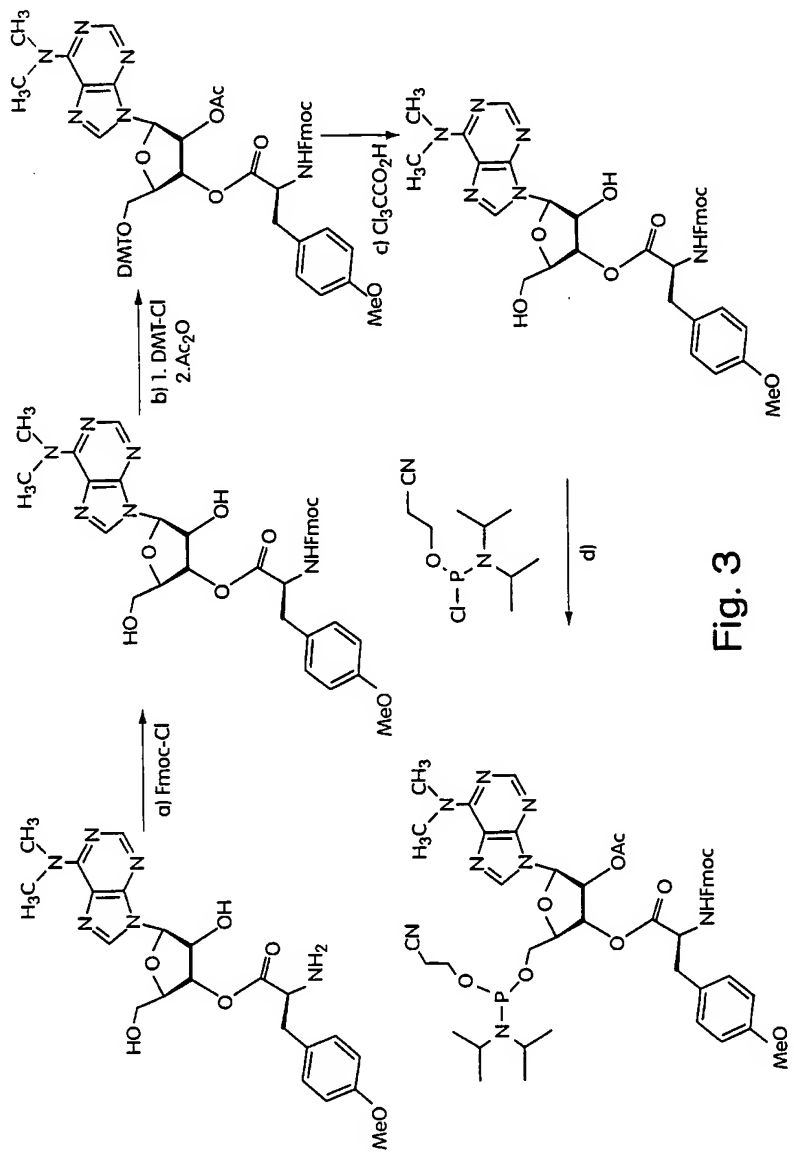

b) 1. DMT-Cl
2.$Ac_2O$
c) $Cl_3CCO_2H$
a) Fmoc-Cl
d)
DMTO
MeO
NHFmoc
OAc
HO
OH
CN

Fig. 3

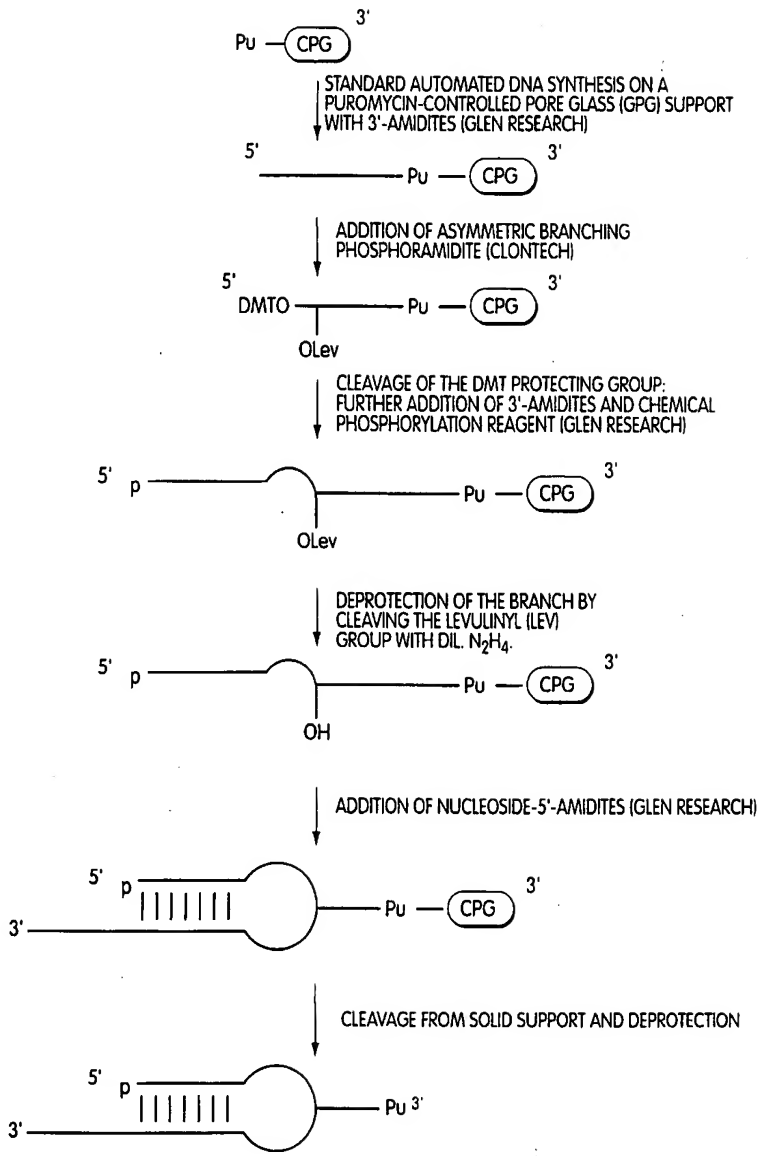

Pu
CPG
3'
STANDARD AUTOMATED DNA SYNTHESIS ON A PUROMYCIN-CONTROLLED PORE GLASS (GPG) SUPPORT WITH 3'-AMIDITES (GLEN RESEARCH)
5'
ADDITION OF ASYMMETRIC BRANCHING PHOSPHORAMIDITE (CLONTECH)
DMTO
OLev
CLEAVAGE OF THE DMT PROTECTING GROUP: FURTHER ADDITION OF 3'-AMIDITES AND CHEMICAL PHOSPHORYLATION REAGENT (GLEN RESEARCH)
p
DEPROTECTION OF THE BRANCH BY CLEAVING THE LEVULINYL (LEV) GROUP WITH DIL. $N_2H_4$.
OH
ADDITION OF NUCLEOSIDE-5'-AMIDITES (GLEN RESEARCH)
CLEAVAGE FROM SOLID SUPPORT AND DEPROTECTION
Pu 3'

Fig. 4







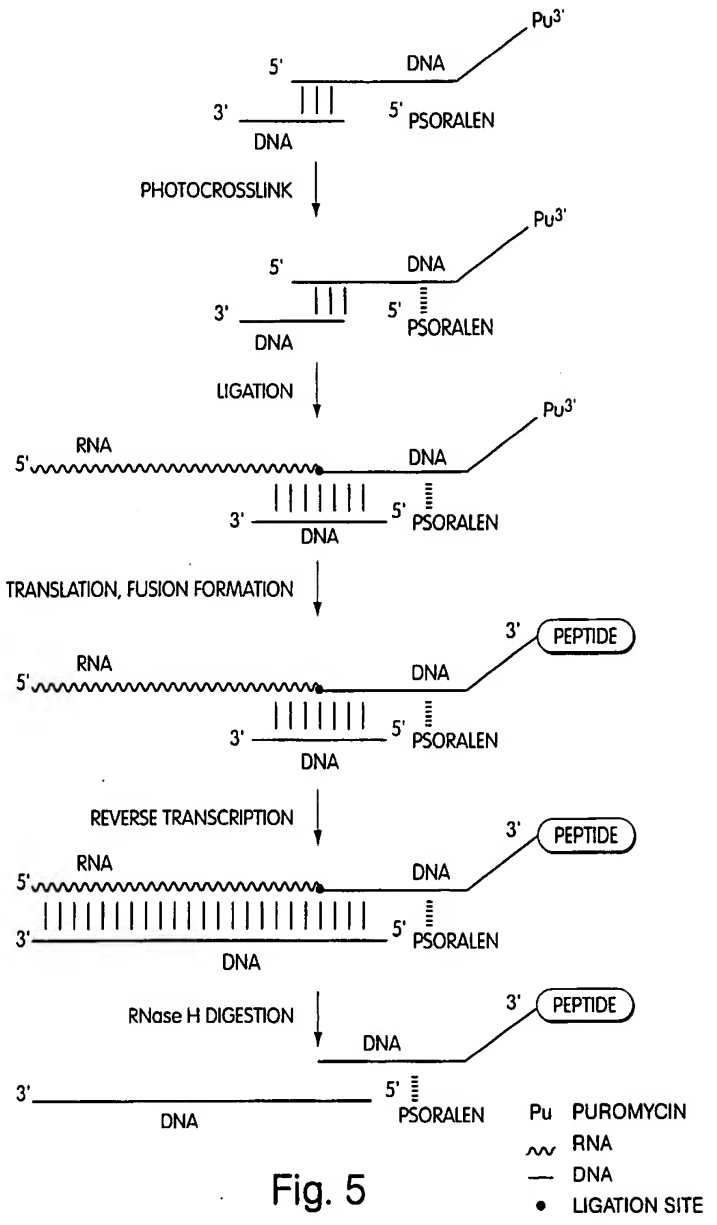

Pu3'
5'
DNA
3'
5'
PSORALEN
DNA
PHOTOCROSSLINK
LIGATION
RNA
TRANSLATION, FUSION FORMATION
PEPTIDE
REVERSE TRANSCRIPTION
RNase H DIGESTION
Pu PUROMYCIN
RNA
DNA
LIGATION SITE

Fig. 5

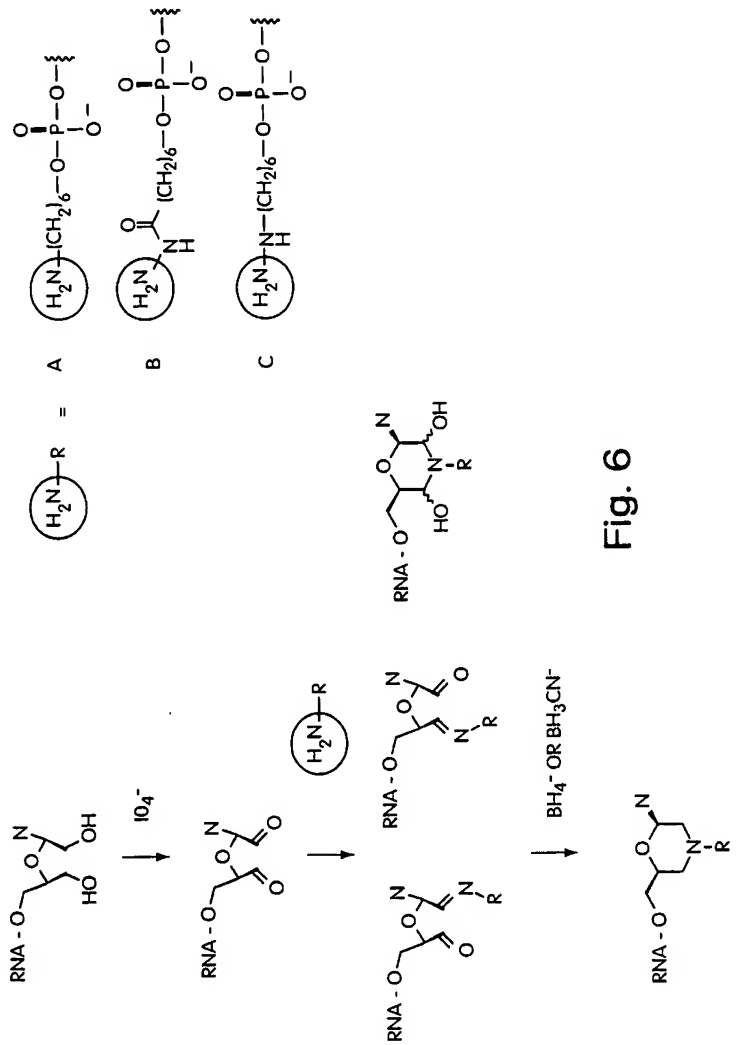

A
B
C
R =
RNA - O
$IO_4^-$
$BH_4^-$ OR $BH_3CN^-$

Fig. 6







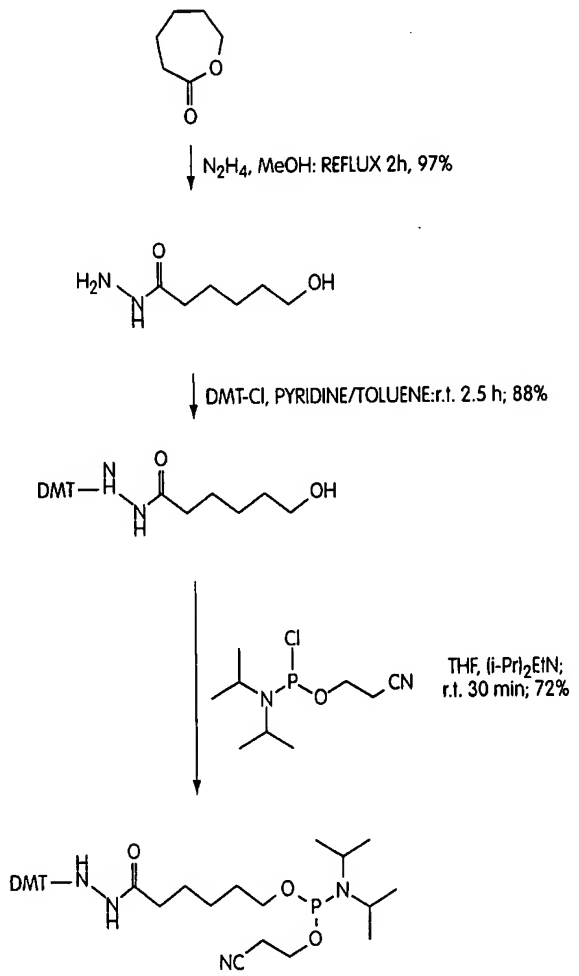

$N_2H_4$, MeOH: REFLUX 2h, 97%
DMT-Cl, PYRIDINE/TOLUENE:r.t. 2.5 h; 88%
THF, $(i\text{-}Pr)_2EtN$;
r.t. 30 min; 72%

Fig. 7

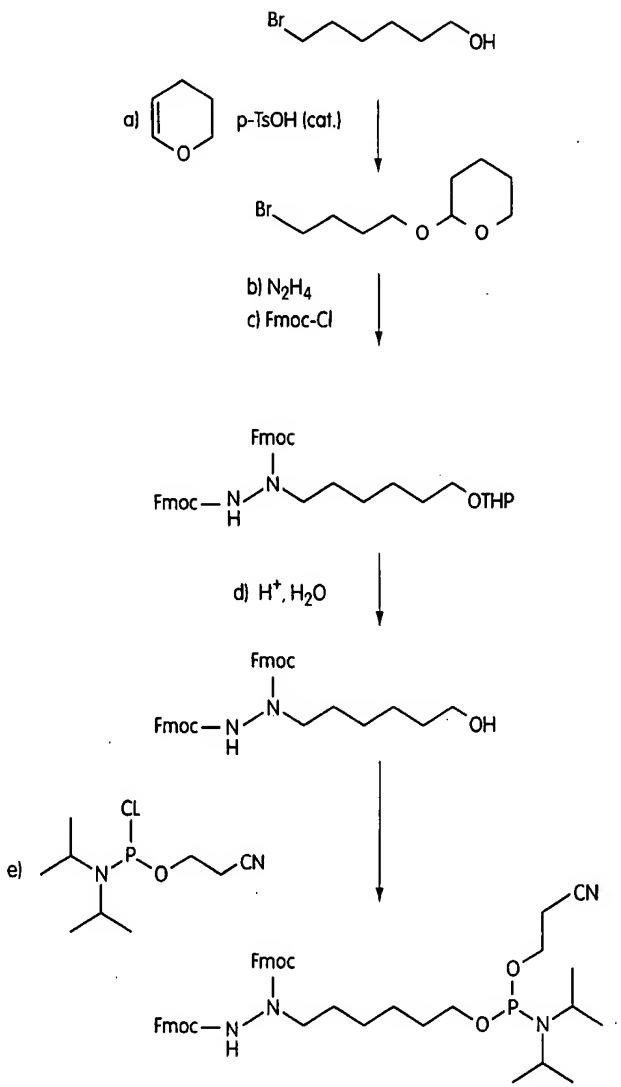

Br
OH
a)
p-TsOH (cat.)
Br
O
O
b) $N_2H_4$
c) Fmoc-Cl
Fmoc
Fmoc—N
N
H
OTHP
d) $H^+$, $H_2O$
Fmoc
Fmoc—N
N
H
OH
e)
CL
P
N
O
CN
CN
Fmoc
Fmoc—N
N
H
O
O
P
N

Fig. 8







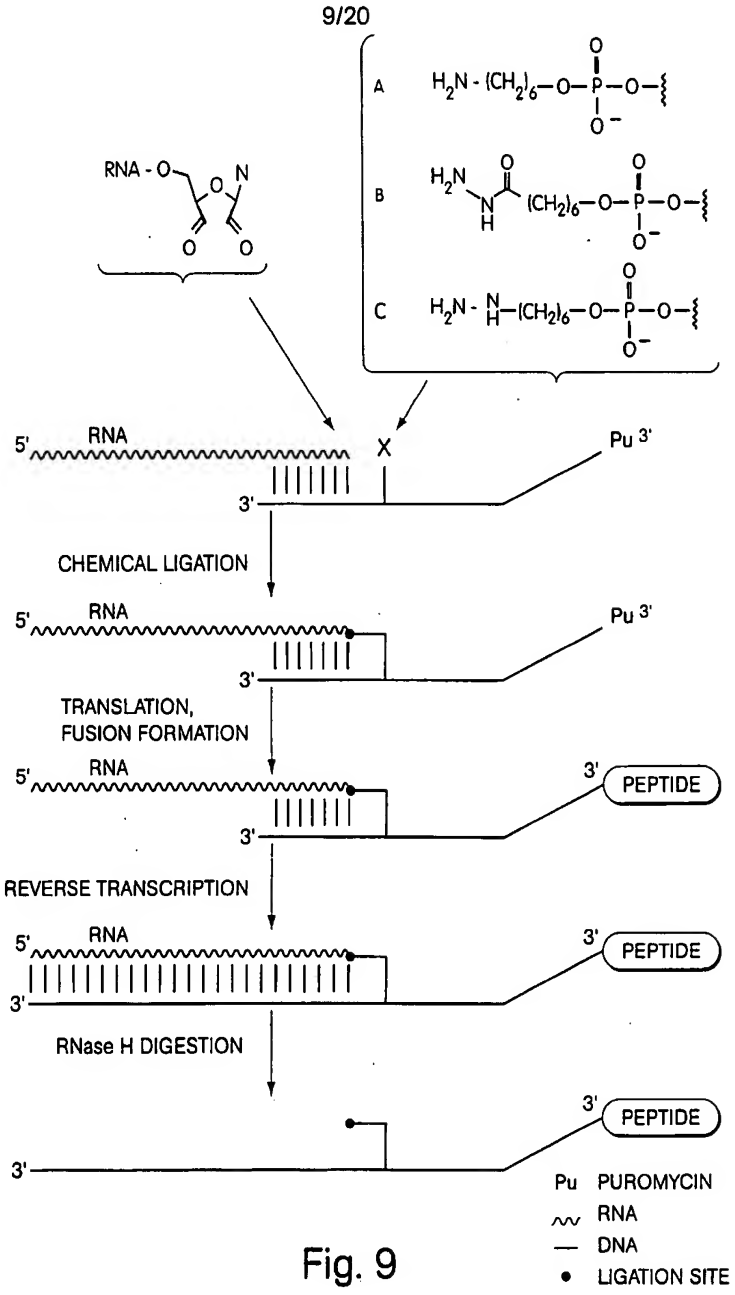

A
$H_2N$ - $(CH_2)_6$—O—P—O—
O⁻
B
$H_2N$
N
H
O
$(CH_2)_6$—O—P—O—
O⁻
C
$H_2N$ - N
H
$(CH_2)_6$—O—P—O—
O⁻
RNA - O
O
N
O
O
5'
RNA
X
Pu 3'
3'
CHEMICAL LIGATION
TRANSLATION,
FUSION FORMATION
PEPTIDE
REVERSE TRANSCRIPTION
RNase H DIGESTION
Pu PUROMYCIN
RNA
DNA
LIGATION SITE

Fig. 9

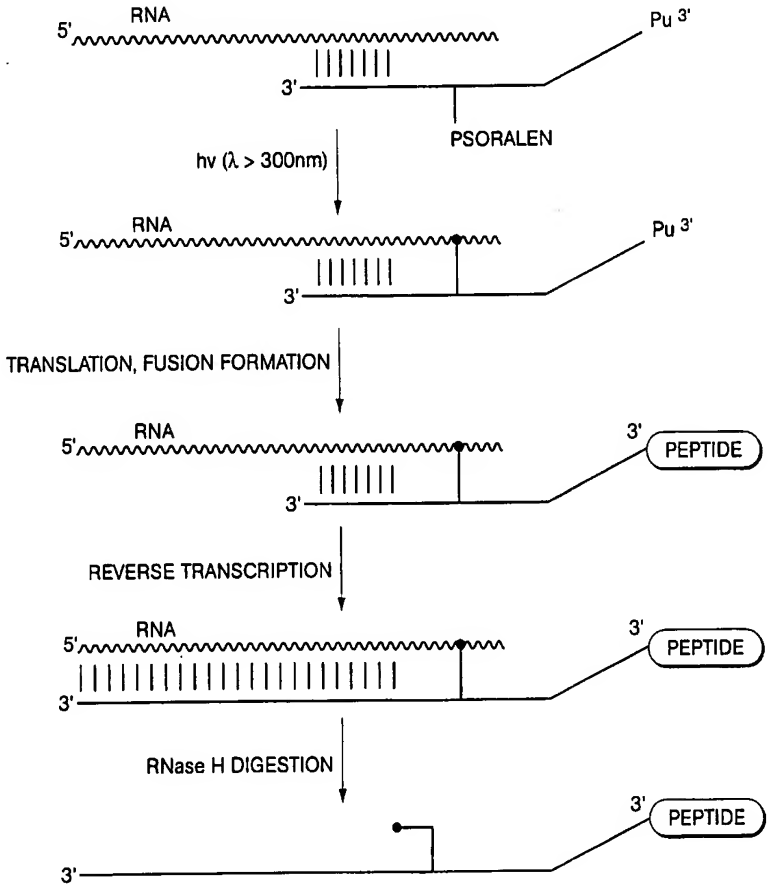



Fig. 10

Pu PUROMYCIN
RNA
DNA
LIGATION SITE





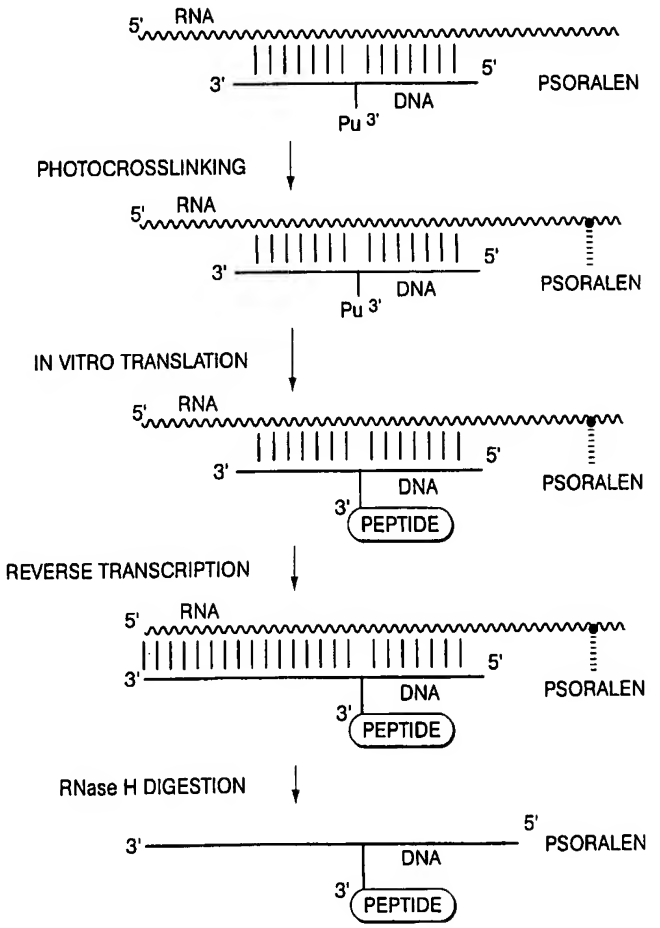



Fig. 11

Pu PUROMYCIN
RNA
DNA
LIGATION SITE

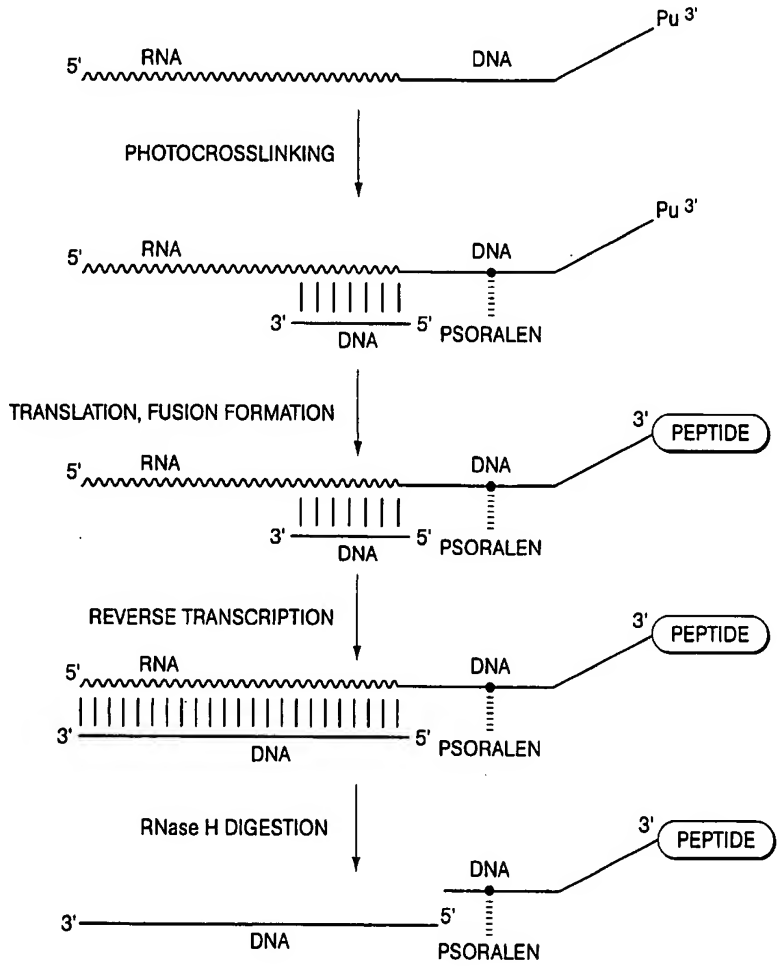



Fig. 12

Pu PUROMYCIN
RNA
DNA
LIGATION SITE







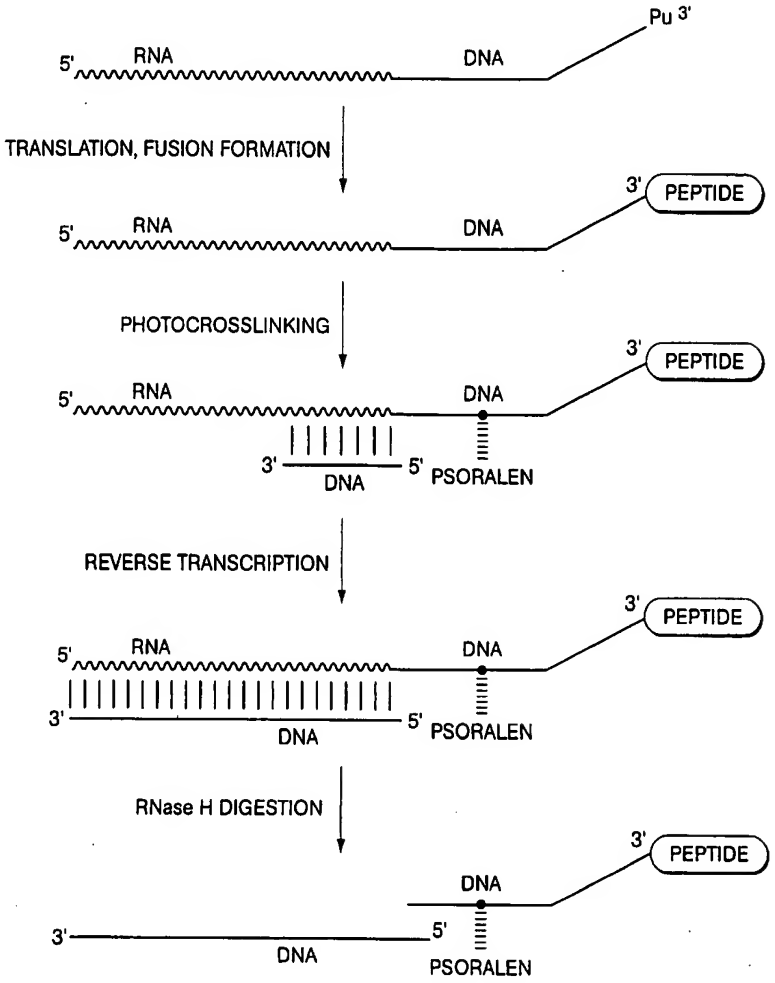



Fig. 13

Pu PUROMYCIN
RNA
DNA
LIGATION SITE

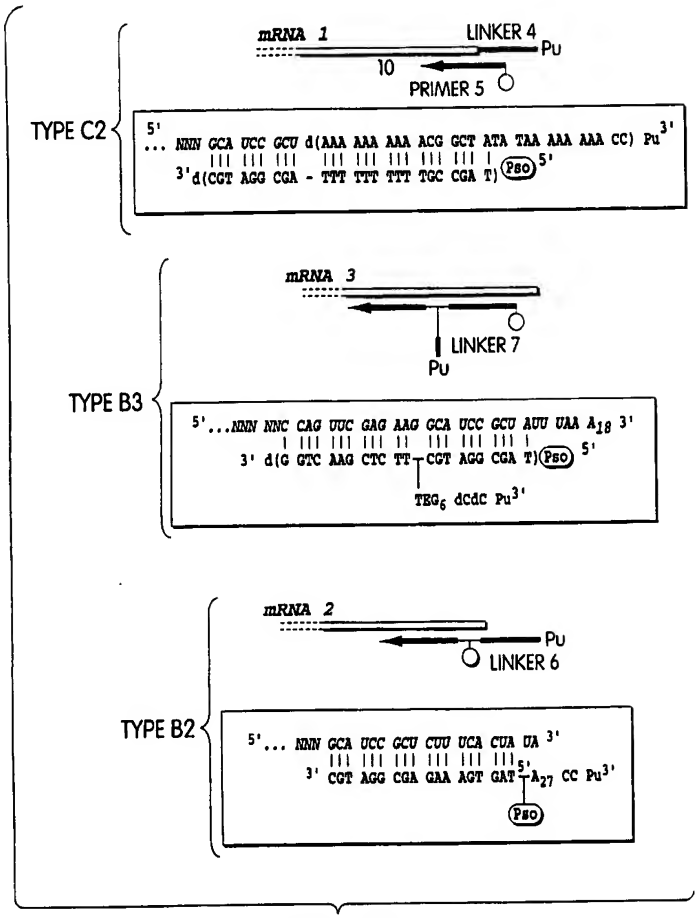

mRNA 1
LINKER 4
Pu
10
PRIMER 5
TYPE C2
5'
... NNN GCA UCC GCU d(AAA AAA AAA ACG GCT ATA TAA AAA AAA CC) Pu 3'
3' d(CGT AGG CGA - TTT TTT TTT TGC CGA T) Pso 5'
mRNA 3
LINKER 7
Pu
TYPE B3
5'...NNN NNC CAG UUC GAG AAG GCA UCC GCU AUU UAA A18 3'
3' d(G GTC AAG CTC TT-CGT AGG CGA T) Pso 5'
TEG6 dCdC Pu 3'
mRNA 2
Pu
LINKER 6
TYPE B2
5'... NNN GCA UCC GCU CUU UCA CUA UA 3'
3' CGT AGG CGA GAA AGT GAT-5'A27 CC Pu 3'
Pso

Fig. 14







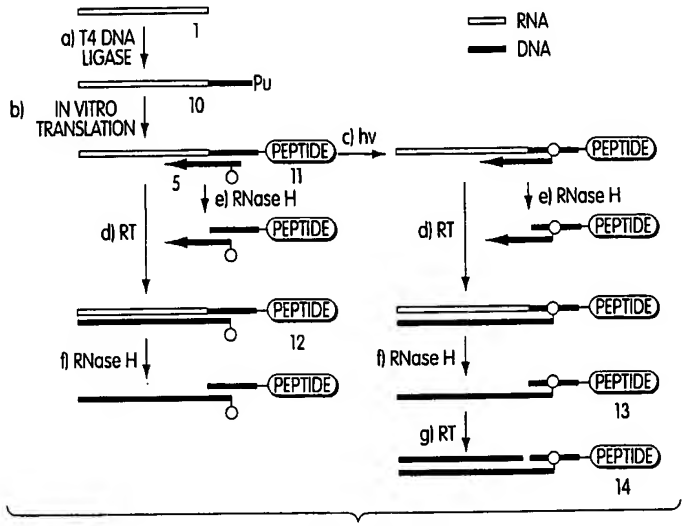

RNA
DNA
a) T4 DNA LIGASE
1
Pu
10
b) IN VITRO TRANSLATION
PEPTIDE
11
c) hv
5
e) RNase H
d) RT
12
f) RNase H
13
g) RT
14

Fig. 15

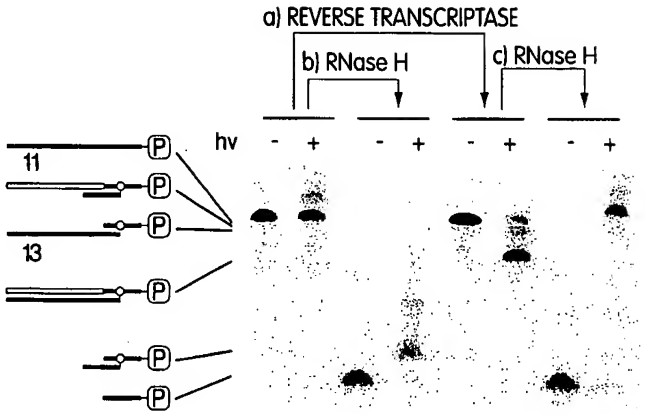

a) REVERSE TRANSCRIPTASE
b) RNase H
c) RNase H
hv
- + - + - + - +
11
13
P

Fig. 16





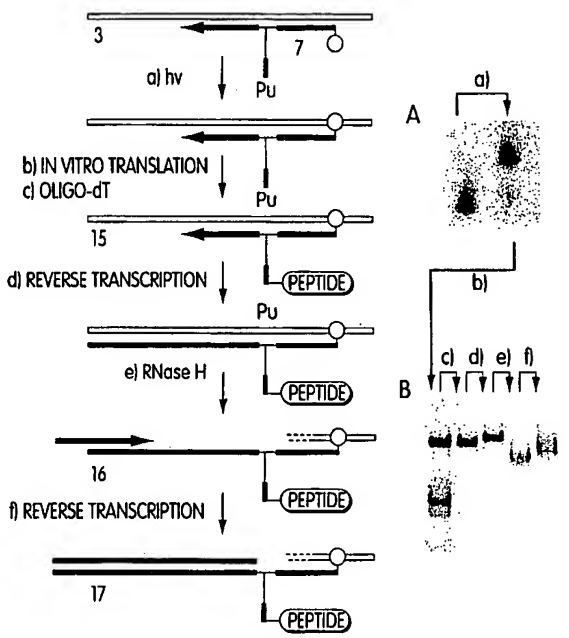

3
7
a) hv
Pu
b) IN VITRO TRANSLATION
c) OLIGO-dT
15
d) REVERSE TRANSCRIPTION
PEPTIDE
e) RNase H
16
f) REVERSE TRANSCRIPTION
17
a)
A
b)
c) d) e) f)
B

Fig. 17

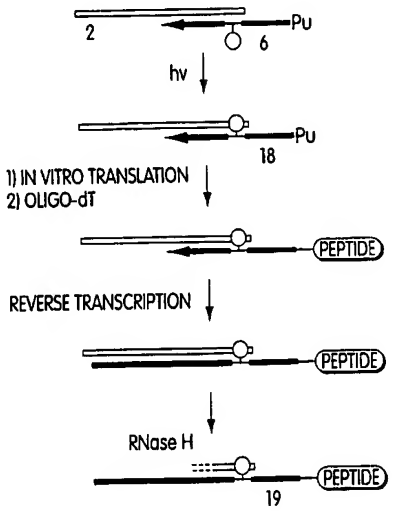

2
6
Pu
hv
Pu
18
1) IN VITRO TRANSLATION
2) OLIGO-dT
PEPTIDE
REVERSE TRANSCRIPTION
PEPTIDE
RNase H
PEPTIDE
19

Fig. 18







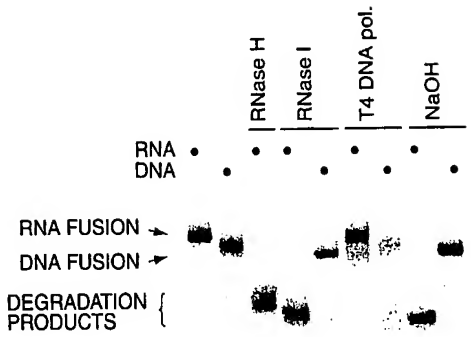

RNase H
RNase I
T4 DNA pol.
NaOH
RNA
DNA
RNA FUSION
DNA FUSION
DEGRADATION PRODUCTS

Fig. 19

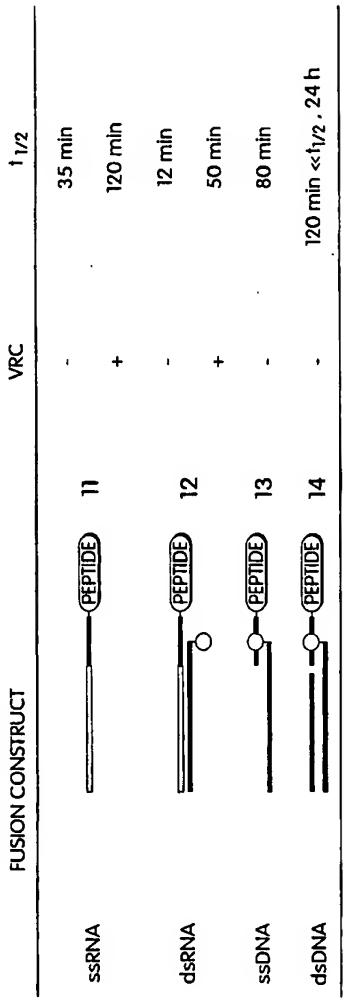



Fig. 20



## INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/28472

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) :C2Q 1/68; C12P 19/34; C07H 21/00, 21/02,21/04
US CL :435/6, 91.1, 91.2, 91.21; 536/22.1, 23.1, 23.4

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1, 91.2, 91.21; 536/22.1, 23.1, 23.4

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST USPATFULL, EPO,JPO,WPIDS; STN-MEDLINE,CAPLUS,EMBASE,BIOSIS, CANCERLIT
search terms: dna-protein fusion covalent,bonding

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 98/31700 A1 (THE GENREAL HOSPITAL CORPORATION) 23 JULY 1998 (23.07.98), see entire document, especially pages 2-3 | 1-54 |
| X | ROBERTS et al. RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. USA. November 1997, vol. 94, pages 12297-12302, see entire document. | 1-54 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"B" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 05 APRIL 2000 | 18 APR 2000 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer<br>JEFFREY SIEW<br>Telephone No. (703) -308-0196 |